DOCUMENT RESUME

ED 436 582 TM 030 444

AUTHOR Johnson, Robert A.; Gerstein, Dean R.
TITLE Potential Sources of Bias in Substance Abuse Treatment Follow-Up Studies.
INSTITUTION National Opinion Research Center, Washington, DC.; Caliber Associates, Fairfax, VA.
SPONS AGENCY Substance Abuse and Mental Health Services Administration (DHHS/PHS), Rockville, MD. Center for Substance Abuse Treatment.
PUB DATE 1999-07-00
NOTE 65p.
CONTRACT 270-97-7016
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Followup Studies; *Outcomes of Treatment; Program Effectiveness; Program Evaluation; *Research Methodology; Responses; *Statistical Bias; *Substance Abuse; Surveys
IDENTIFIERS Randomization (Statistics)

ABSTRACT

Large multi-site followup studies of treatment clients comprise a major source of evidence on the effectiveness of substance abuse treatment in the United States. This report considers challenges to the validity of results in treatment effectiveness studies due to four factors: (1) nonrandom entry to treatment from the population needing treatment; (2) nonrandom selection of treatment providers by researchers for follow-up studies; (3) non-cooperation of the selected treatment providers with the research protocol; and (4) incomplete enrollment and follow-up of clients treated by the cooperating providers. Both kinds of selection/entry and both kinds of non-response are potential sources of bias, so both should be major concerns in the design and analysis of substance abuse treatment follow-up surveys. The response characteristics of four large-scale multi-site studies carried out in the early and mid 1990s are summarized and assessed on a comparative basis. Recommendations are provided for minimizing bias in future studies, and statistical methods are proposed for evaluating biases due to the process of entry into treatment and non-response. An appendix describes the National Treatment Improvement Evaluation Study and Center for Substance Abuse Treatment Demonstrations. (Contains 7 exhibits and 52 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 436 582

NEDS

NATIONAL EVALUATION DATA SERVICES

POTENTIAL SOURCES OF BIAS IN SUBSTANCE ABUSE TREATMENT FOLLOW-UP STUDIES

July 1999

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
R. Finkbiner
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

National Opinion Research Center

BEST COPY AVAILABLE

TM030444

NEDS

NATIONAL EVALUATION DATA SERVICES

POTENTIAL SOURCES OF BIAS IN SUBSTANCE ABUSE TREATMENT FOLLOW-UP STUDIES

Prepared by

Robert A. Johnson, Ph.D.
Dean R. Gerstein, Ph.D.

National Opinion Research Center
1350 Connecticut Avenue, NW, Suite 500
Washington, DC 20036

July 1999

BEST COPY AVAILABLE

This document was supported by the Center for Substance Abuse Treatment, Department of Health and Human Services, Caliber/NEDS Contract No. 270-97-7016.

CSAT
Center for Substance Abuse Treatment
SAMHSA

TABLE OF CONTENTS

Page

FOREWORD i

ACKNOWLEDGMENTS ii

ABSTRACT iii

I. INTRODUCTION 1

1. ABSENCE OF RANDOM ASSIGNMENT TO TREATMENT 1

2. NONRESPONSE OCCURRING AT EACH OF TWO STAGES OF SAMPLE SELECTION 3

3. ORGANIZATION OF THIS REPORT 3

II. OVERVIEW OF NONRESPONSE IN FOUR SUBSTANCE ABUSE TREATMENT FOLLOW-UP STUDIES 5

1. CALDATA 11

2. SROS 12

3. NTIES 13

4. DATOS 15

5. SUMMARY 17

III. CALDATA: NONRESPONSE AT THE PROVIDER AND CLIENT STAGES OF SAMPLE SELECTION 19

1. DESIGN FEATURES CONTRIBUTING TO HIGHER RESPONSE RATES 19

2. LESSONS FROM CALDATA 23

TABLE OF CONTENTS (CONTINUED)

Page

3. NONRESPONSE BIAS 24

3.1 First Stage—Bias Due to Provider Noncooperation 24
3.2 Second Stage—Bias Due to Client Nonresponse in Cooperating Providers 25

IV. BIAS DUE TO SELECTION INTO TREATMENT 28

V. METHODS FOR ESTIMATING AND EVALUATING SELECTION BIAS 33

1. WEIGHTING ADJUSTMENTS INCLUDING POSTSTRATIFICATION 35

2. LIKELIHOOD-BASED METHODS 36

3. IMPUTATION 38

VI. IMPLICATIONS FOR TREATMENT RESEARCH, POLICY, AND PRACTICE 41

REFERENCES 44

APPENDIX: DESCRIPTION OF THE NATIONAL TREATMENT IMPROVEMENT EVALUATION STUDY AND CENTER FOR SUBSTANCE ABUSE TREATMENT DEMONSTRATIONS (1990-1992)

FOREWORD

The Center for Substance Abuse Treatment (CSAT) works to improve the lives of those affected by alcohol and other substance abuse, and, through treatment, to reduce the ill effects of substance abuse on individuals, families, communities, and society at large. Thus, one important mission of CSAT is to expand the availability of effective substance abuse treatment and recovery services. To aid in accomplishing that mission, CSAT has invested and continues to invest significant resources in the development and acquisition of high-quality data about substance abuse treatment services, clients, and outcomes. Sound scientific analysis of this data provides evidence upon which to base answers to questions about what kinds of treatment work best for what groups of clients, and about which treatment approaches are cost-effective methods for curbing addiction and addiction-related behaviors.

In support of these efforts, the Program Evaluation Branch (PEB) of CSAT established the National Evaluation Data Services (NEDS) contract to provide a wide array of data management and scientific support services across various programmatic and evaluation activities. Essentially, NEDS is a pioneering effort for CSAT in that the Center previously had no mechanisms established to pull together databases for broad analytic purposes or to house databases produced under a wide array of activities. One of the specific objectives of the NEDS project is to provide CSAT with a flexible analytic capability to use existing data to address policy-relevant questions about substance abuse treatment. This report has been produced in pursuit of this objective.

This report explores two methodological issues of importance to substance abuse treatment researchers and policy analysts alike—nonresponse by members of a cohort of treatment clients whose behavior is being studied across time, and selection bias in recruiting such cohorts to begin with. The purpose of the analyses being reported here is to consider the extent and effects of nonresponse and selection bias in a recent series of four large-scale follow-up studies: the California Drug and Alcohol Treatment Assessment (CALDATA); the Services Research Outcomes Study (SROS); the National Treatment Improvement Evaluation Study (NTIES); and the Drug Abuse Treatment Outcome Study (DATOS). The report also includes some suggested approaches for evaluating the robustness of conclusions, considering these potential threats to the validity of results.

Sharon Bishop
Project Director
National Evaluation Data Services

ACKNOWLEDGMENTS

Caliber Associates is the prime contractor for NEDS in partnership with Battelle Centers for Public Health Research and Evaluation (CPHRE); the Lewin Group; National Opinion Research Corporation (NORC) and Computech, Inc. We wish to thank Robert A. Johnson, Ph.D., and Dean R. Gerstein, Ph.D., of the National Opinion Research Center, who authored this report. Thanks are also due to Substance Abuse and Mental Health Services Administration (SAMHSA) staff members who reviewed and commented on an earlier draft of this paper. We wish to acknowledge the guidance and direction of Ron Smith, Ph.D., Government Project Officer for the NEDS contract. In addition, the following individuals within the NEDS staff made significant contributions to the production of this document: Marsha Morahan, Dana Vaughn, and Kim Nguyen.

ABSTRACT

Large multisite follow-up studies of treatment clients comprise a major source of evidence on the effectiveness of substance abuse treatment in the United States. This report considers challenges to the validity of results in treatment effectiveness studies due to four factors: nonrandom entry to treatment from the population needing treatment; nonrandom selection of treatment providers by researchers for follow-up studies; noncooperation of the selected treatment providers with the research protocol; and incomplete enrollment and follow-up of clients treated by the cooperating providers. Both kinds of selection/entry and both kinds of nonresponse are potential sources of bias, so both should be major concerns in the design and analysis of substance abuse treatment follow-up surveys. We summarize and assess on a comparative basis the response characteristics of four large-scale multisite studies carried out in the early and mid 1990s, provide recommendations for minimizing nonresponse in future studies, and propose statistical methods for evaluating biases due to the process of entry into treatment as well as to nonresponse.

I. INTRODUCTION

A number of large-scale multi-site observational follow-up studies of substance abuse treatment performed during recent decades have been instrumental in persuading many researchers and policy analysts that substance abuse treatment programs in the U.S. are highly cost-effective (see, for example, Office for National Drug Control Policy, 1998). As the investigators leading these studies have readily acknowledged, observational methods are not the optimal way to precisely measure treatment effects; these studies are, however, practical in ways that large-scale randomized clinical trials are not.

Follow-up studies involve methodological questions that merit more attention than they have yet received. Nonresponse by members of a panel of treatment clients whose behavior is being studied across time is one such important methodological issue; selection bias in recruiting such panels to begin with is another such issue. The purpose of the following analyses is to consider the extent and effects of nonresponse and selection bias in a recent series of four large-scale follow-up studies, to evaluate the potential vitiating effects of nonresponse and selection bias, and to suggest some approaches for testing the robustness of conclusions in the face of potential threats to the validity of results.

From the standpoint of possible selection biases, major substance abuse treatment follow-up studies in the U.S. share two important methodological features: absence of random assignment to treatment and nonresponse occurring at each of two stages of sample selection.

1. ABSENCE OF RANDOM ASSIGNMENT TO TREATMENT

Clients enrolled in substance abuse treatment follow-up studies are generally sampled from the universe of individuals admitted to or discharged from treatment during a specified time frame rather than from the universe of individuals in need of treatment or potentially benefitting from treatment services. The subset of individuals who enter treatment arises from a mutual process of selection by potential clients and substance abuse treatment programs: Individuals choose to apply for treatment, and programs choose which applicants to accept. Some individuals enter treatment "voluntarily" while others are pressed to seek treatment by the criminal justice system, as an alternative to incarceration or extended supervision. In most cases, individuals are induced to enter treatment by a variety of factors: subjective motivations such as depression, guilt, fear, or the pain of illness; pressures from families, friends, employers, police, and others; changes in local markets for preferred substances; and perceptions about whether

treatment is of good quality, accessible, and affordable (Gerstein and Harwood, 1990). None of these factors are well measured in the general population of substance users.

In household surveys of the U.S. population, it is difficult to identify a nontreated sample of substance-abusing individuals not receiving treatment who are comparable to those in treatment. There are several reasons. First, many individuals in treatment do not reside in households and thus are not represented in household surveys. Second, there exists no updated master list of individuals with severe substance abuse problems that could be used as a sampling frame for developing a nontreatment control group. Third, programs make decisions to accept applicants based on multiple factors, including payment resources, specific exclusion or preference criteria, and capacity controls that may be specific to programs and locales that may be difficult to measure and control in developing a nontreatment control group.

Nevertheless, several considerations suggest that the appropriate target population for inferences from substance abuse treatment follow-up studies is the population in need of treatment rather than the population subset actually admitted to (or discharged from) treatment during the reference period. These considerations include the following:

- Many inferences are appropriately framed with respect to substance-abusing individuals not in treatment. For example, policies designed to affect treatment effectiveness may also affect the process and probability of initial or repeated entry into treatment. Evaluation of policy initiatives often focuses on unadmitted populations in need of treatment as a whole or in terms of special population segments that are "underserved."

- Selection into treatment is contingent on factors mentioned above that may vary in time and across localities. Since substance abuse treatment follow-up surveys differ according to time frame and sampled localities, comparing survey results requires taking the selection process into account.

- As explained later, the internal validity of analytical conclusions in substance abuse treatment follow-up studies may be compromised by features of the entry process even if the scope of conclusions is explicitly restricted to the treatment population.

Aside from its potential importance in understanding the effectiveness of substance abuse treatment, the analysis of successive movements of individuals into and out of substance abuse treatment programs—sometimes referred to as "treatment careers"—is an important topic in its own right. Episodes of treatment tend to be of short duration relative to the span of substance use careers. The median duration of treatment episodes ranges from a few weeks to a year

depending on the type of treatment. In many treatment programs, the majority of patients are in their second, third, or later episode. There are high rates of mobility between the in-treatment population and the nontreatment population that is in need of treatment.

2. NONRESPONSE OCCURRING AT EACH OF TWO STAGES OF SAMPLE SELECTION

Major substance abuse treatment follow-up surveys in the U.S. have employed two-stage sample designs. The first stage samples treatment providers,[1] and the second stage samples clients within the selected providers. These surveys have evaluated treatment effectiveness by means of follow-up interviews with sampled clients after they have left treatment. Nonresponse occurs at the first stage due to noncooperation of the sample providers. Nonresponse occurs at the second stage due to problems of locating and obtaining interviews from sampled clients selected from cooperating providers. Estimates of treatment effectiveness based on these surveys are biased to the extent that first and second stage nonresponse rates are large and to the extent that nonresponding providers and clients are different from responding providers and clients.[2]

3. ORGANIZATION OF THIS REPORT

The second section of this report provides an overview of selection and nonresponse issues in four major U.S. treatment follow-up studies conducted during the 1990s. Reports published from each of these studies compared responding and nonresponding sample clients to evaluate second-stage nonresponse bias, i.e., bias due to inability to locate and refusals to be interviewed of clients sampled from cooperating providers. One of the four CALDATA reports was able to assess first-stage nonresponse bias, i.e., bias due to the noncooperation of sample providers. None of the four studies was able to assess possible biases due to nonrandom selection into treatment.

---

1 Each "provider" typically comprises one or more service delivery unit (SDU) specializing in particular types of treatment ("modalities of treatment"), such as residential treatment, methadone maintenance, and nonmethadone outpatient treatment.

2 If p_r is an estimate of treatment effectiveness (e.g., the percentage of clients who are drug-free 1 year after leaving treatment) calculated using data from survey respondents, and p_{true} is the true (unknown) value of the same measure, then the bias due to nonresponse equals $(p_r - p_{true}) = NR \cdot (p_r - p_{nr})$, where NR is the nonresponse rate and p_{nr} denotes the same measure of treatment effectiveness among nonrespondents (Groves, 1989, p. 133).

The third section presents a detailed secondary analysis evaluating first-stage and second-stage nonresponse bias associated with measured covariates in CALDATA. The fourth section discusses selection bias due to the process of entry into treatment. The fifth section discusses methods for estimating and evaluating biases due to both nonresponse and selection into treatment, even when it is not feasible to select a control or comparison group. The sixth section summarizes our conclusions and recommendations for further research.

II. Overview of Nonresponse in Four Substance Abuse Treatment Follow-Up Studies

Exhibit II-1 summarizes the target populations, sampling methods, research designs, and sample sizes of four recent U.S. substance abuse treatment follow-up surveys:

- CALDATA—California Drug and Alcohol Treatment Assessment (Gerstein et al., 1994): a panel of 3,055 clients selected from client records abstracted in 87 clinical units, followed up an average of 15 months after discharge from treatment
- SROS—Services Research Outcomes Study (Schildhaus et al., 1998): a panel of 3,047 clients selected from client records abstracted in 99 units, followed up about 5½ years after discharge from treatment
- NTIES—National Treatment Improvement Evaluation Study (Gerstein, Datta et al., 1997): a panel of 6,593 clients completing intake interviews in 71 units followed up an average of nearly 1 year after discharge from treatment
- DATOS—Drug Abuse Treatment Outcome Study first-year follow-up (Simpson, & Curry, 1997): a panel of 4,786 clients completing intake interviews in 76 units, followed up an average of 1 year after discharge from treatment.

The last three rows of Exhibit II-1 present variously calculated response rates of the four surveys. Each of the three sets of response rates takes into account only "client nonresponse" or "second-stage nonresponse," i.e., nonresponse due to inability to locate for follow-up or refusals to be interviewed by clients selected for follow-up from cooperating providers included in the follow-up phase. The response rates do *not* take into account first-stage nonresponse, i.e., nonresponse due to noncooperating sample providers and other providers excluded from the follow-up.

The last three rows of Exhibit II-1 represent an attempt to transform the measures, which are prepared somewhat differently in the published reports of each study, to a series of equivalent bases. The first of these rows indicates the percentage of the selected follow-up panel actually interviewed during the follow-up period, ranging from 59 percent of the SROS panel to 82 percent of the NTIES panel. The second row eliminates from this response rate calculation the sampled persons known to have been deceased during the interval between the impanelment interval and completion of the follow-up fieldwork period. Since mortality rates were in the range of 1-2 percent of sample per year, this adjustment provides a somewhat more realistic assessment of follow-up effectiveness encompassing time periods of different durations. Both calculations indicated that SROS, CALDATA, and DATOS were nearly equivalent in their effectiveness in obtaining follow-up interviews, gaining data from nearly four out of every six

EXHIBIT II-1
FOUR RECENT FOLLOW-UP SURVEYS OF DRUG AND ALCOHOL TREATMENT CLIENTS IN THE U.S.[1]

	CALDATA	SROS	NTIES	DATOS
Sponsor	State of California/ADP	SAMHSA/Office of Applied Science	SAMHSA/CSAT	NIDA
Survey performed by	National Opinion Research Center (NORC)	National Opinion Research Center (NORC)	NORC and Research Triangle Institute (RTI)	Research Triangle Institute (RTI)
Target population of substance abuse treatment clients	Discharged from publicly-funded California SDUs, 10/91- 9/92	Discharged 9/89-8/90 from a U.S. drug treatment facility	Admitted to SDU included in CSAT demonstration project, 7/93- 9/94	Treated in a "typical and stable" SDU in one of 11 U.S. cities, 11/91-Fall/93
When follow-up was fielded	4/1993-12/1993	6/1995-3/1996	8/1993-10/1995	1994-1996
Average follow-up period after discharge (range)	15 months (9-24)	5 years (58-74 months)	11 months (5-20)	1 year
Repeated measurements: pre-/during/post-treatment	SDU records abstracted; follow-up interviews only	SDU records abstracted; follow-up interviews only	Intake, discharge, and follow-up interviews	Intake, in-treatment, and follow-up interviews
Sampling of SDUs	Probability sampling stratified by modality and geography, in California	Probability sampling stratified by modality and geography, among 48 states	Purposive sampling: 16 cities, based on modality, grant focus, cooperation	Purposive sampling: 11 cities, partial replicate of TOPS sites
Modalities of treatment (original terms from each study)	Social model Other residential Methadone detoxification Methadone maintenance Outpatient nonmethadone	Hospital inpatient Residential Outpatient methadone Outpatient drug-free	Short-term residential Long-term resid (>2 mos) Outpatient methadone Outpatient nonmethadone Correctional facilities	Short-term inpatient Long-term resid (>6 mos) Outpatient methadone Outpatient drug-free
# Sampled/Selected SDUs	106	146	82	120
# Cooperating SDUs	87	99	71	96
# Represented SDUs	87	99	71	76
# Clients in follow-up panel	3,055	3,047	6,593	4,786
# Follow-up interviews	1,858	1,799	5,388	2,966

BEST COPY AVAILABLE

EXHIBIT II-1 (CONTINUED)
FOUR RECENT FOLLOW-UP SURVEYS OF DRUG AND ALCOHOL TREATMENT CLIENTS IN THE U.S.[1]

	CALDATA	SROS	NTIES	DATOS
Panel follow-up response rate[2]	61%	59%	82%	62%
Panel follow-up response rate, excluding deceased[3]	64%	65%	82%	63%
Response rate for all available clients, excluding deceased[4]	64%	65%	70%	48%

1. Details are presented in Gerstein et al. (1994), Schildhaus et al. (1998), Gerstein, Datta et al. (1997), and Simpson and Curry (1997), respectively.
2. The panel follow-up rates pertain to completed interviews among sample clients in cooperating SDUs enrolled successfully in the pertinent study panel; that is, clients who completed an intake interview (two intake interviews in the case of DATOS, which also subsampled to select the group participants in the follow-up study; see text and Flynn et al. in Simpson and Curry, 1997). These rates do not take into account eligible/recruited but non-enrolled clients e.g., through refusal to participate in intake interviews) or any clients in noncooperating sample SDUs. Also, the response rates do not take into account possible undercoverage in SDU records of eligible clients.
3. Sample individuals known to be deceased at the time of the follow-up interview are excluded from the numerator and the denominator. Other panel cases ultimately classified by authors as ineligible for follow-up are included in this denominator; these cases were not followed up for reasons such as language difficulties (neither English nor Spanish), incapacitating illness, residence at the time of follow-up was too far from the SDU site, or location in relatively inaccessible hospitals, jails, or prisons. The specific rules and extent of these exclusions varied, and the percentage classified as ineligible for these reasons ranged from fewer than 1 percent in SROS to 12 percent in DATOS.
4. The denominator in these calculations is all clients receiving treatment in cooperating SDUs during the period of study induction, less those excluded due to strictly random sampling and less those known to be deceased. For CALDATA and SROS, this denominator is the same as the selected panel (less deceased). For NTIES, the denominator is the aggregate of all eligible patients—those receiving one or more units of treatment—recorded in cooperating SDUs, for whom an 85 percent research induction rate (research intake interviews completed per all eligible) was achieved; less the deceased. The DATOS response rate is an estimate. Admissions records were not centrally collated as in NTIES, so the induction rate for the initial intake interview cannot be precisely determined with available data. We have estimated this induction rate at 90 percent, but it may have been higher (95% would be practical maximum) or lower (if comparable to NTIES, 85%). In addition, the 8,109 panel members eligible for follow-up sample selection are based on successful completion of a second intake interview. This number excludes 20 SDUs with less than 20 such completions. If we reclassify these 20 as noncooperating providers, the net completion rate of second intake interviews is approximately 84 percent. Therefore the estimated DATOS rate is approximately (63%)(90%)(84%)=48%.

BEST COPY AVAILABLE

living panel members; and that NTIES yielded an appreciably higher rate, completing interviews with five of every six living panel members. There were substantial differences between studies in resource expenditures for these follow-up interviews, as discussed below.

The final row compares follow-up rates in a larger framework, namely in terms of the total treated population potentially available for study in each cooperating clinic. This measure improved comparability by adjusting for the two different ways in which the panels were initially drawn. Unlike the CALDATA and SROS panels, which were targeted for follow-up without any requirement of previous research interviewing, NTIES and DATOS cases were followed up only if they had been successfully interviewed shortly after admission to treatment (one intake interview was required in NTIES, two were required in DATOS). Numerating the follow-up yields of NTIES and DATOS over these larger groups of eligible clients makes the CALDATA, SROS, and NTIES response look much more comparable, all of them falling within a response rate range of 64-70 percent. However, the DATOS yield drops below 50 percent.

Exhibit II-2 shows that, although the four surveys differed in their allocation of sample cases to modalities of treatment, the client or second-stage follow-up response rates of the four surveys follow approximately the same order—with NTIES highest and the others roughly comparable—when response rates are compared within treatment modalities.[3] This suggests that aspects of research design and implementation other than sample allocation to modalities are responsible for the overall differences among surveys in second-stage response rates.

The final column of Exhibit II-2 shows sample percentages by modality based on the 1995 One Day National Census of the Uniform Facility Data Set, or UFDS (SAMHSA, 1997). The UFDS is intended to be a census, or 100 percent sample, of specialty substance abuse treatment facilities in the U.S. The Substance Abuse and Mental Health Services Administration (SAMHSA) has recognized some limitations of UFDS data, including possible problems of coverage of the full target universe and non-response by some facilities. Nevertheless, the UFDS sample percentages in the final column can serve as a baseline for comparing the results from the

---

[3] Modalities were defined differently in the four surveys. The main discrepancies pertain to "short-term residential," which means a residential treatment program with a typical duration or planned length of stay of less than 6 months in DATOS, less than 2 months in NTIES, and 30 days or less in the Uniform Facility Data Set (UFDS) (final column of Exhibit II-2). In SROS, "short-term residential" means a residential treatment program located in a hospital setting. In CALDATA, "short-term residential" means a particular variety of short-term non-hospital residential treatment, called "the California social model." "Correctional" programs—included only in NTIES—encompass all kinds of treatment programs located in Federal and non-Federal correctional facilities.

EXHIBIT II-2
CLIENT NONRESPONSE RATES AND SAMPLE PERCENTAGES BY MODALITY OF TREATMENT.[1]
CALDATA, SROS, NTIES, DATOS, AND UFDS

	CALDATA		SROS		NTIES		DATOS		UFDS[2]
	Response	Sample %	Response	Sample %	Response	Sample %	Response	Sample %	Sample %
Total sample	61%	100%	59%	100%	82%	100%	62%	100%	100%
Short-Term Residential	56%	24%	58%	25%	79%	21%	63%	27%	3.5%
Long-Term Residential	57%	21%	60%	26%	82%	17%	56%	25%	8.6%
Outpatient, Methadone	66%	32%	58%	24%	87%	8%	66%	24%	12%
Detoxification	58%	18%	–	–	–	–	–	–	1%
Maintenance	77%	14%	–	–	–	–	–	–	11%
Outpatient, Nonmethadone	61%	24%	61%	26%	81%	31%	62%	24%	76%
Correctional	–	–	–	–	84%	23%	–	–	–

1. Response rates pertain to completed interviews among sample clients in cooperating SDUs and do not take into account noncooperating sample SDUs.
2. Uniform Facility Data Set, 1995 One Day National Census. See Substance Abuse and Mental Health Services Administration (1997a, Table 10).

BEST COPY AVAILABLE

four substance abuse treatment follow-up surveys.[4] All four treatment follow-up surveys oversampled clients in short-term and long-term residential programs, and undersampled clients in outpatient programs, relative to the One Day National Census. Since response rates tend to be slightly lower in residential than in outpatient programs, post-stratification using UFDS would slightly raise the overall response rates of each survey.[5]

The reasons for the neglect of first-stage nonresponse in published reports of the four treatment follow-up surveys—except in CALDATA (see below)—are instructive: Two of the four surveys, NTIES and DATOS, used purposive sampling, rather than probability sampling, of treatment providers and their discrete SDUs in the first stage. That is, the treatment providers and SDUs were chosen by the study designers rather than randomly selected with known probabilities from the target population of providers and SDUs. This implies that it is neither meaningful nor possible to use either NTIES or DATOS to estimate first-stage nonresponse in any well-defined target population of treatment providers (see, e.g., Kish, 1965; Cochran, 1972).

The other two surveys, CALDATA and SROS, did use probability sampling. That is, CALDATA and SROS gave every provider in a well-specified target population a known probability of being selected into the sample. Even so, the estimation of first-stage nonresponse is problematic in SROS, because the SROS first-stage sample of providers was selected from cooperating providers in a prior survey conducted during 1989-1990, the Drug Services Research Survey (DSRS), and because published information about noncooperating providers in DSRS is somewhat incomplete.[6] Of the four major recent surveys, only CALDATA currently provides

---

[4] Comparisons between the follow-up surveys and UFDS are not exact, because the numbers of clients in the follow-up surveys are based not on a one-day census but rather on discharges in CALDATA and SROS and on admissions in NTIES and DATOS (see Exhibit II-1). Since the 1995 UFDS was completed, SAMHSA has added about 3,000 substance abuse treatment facilities to the national sampling frame, increasing the total number of listed facilities by about 28 percent. Yet the 1995 UFDS might better approximate the universe of facilities represented by surveys conducted in the early 1990s.

[5] Post-stratification means that data for sample clients in the follow-up study would be classified according to variables measured in both the follow-up study and UFDS, including modality, and weighted so that the percentage distributions of these variables in the follow-up study matches their distribution in UFDS. As discussed in section 5, post-stratification could also be used to reduce bias and increase precision of estimated treatment effects.

[6] Table 2-1 of Schildhaus et al. (1998) estimates that 67 percent of eligible facilities that were sampled in DSRS cooperated and uses this estimate together with the SROS second-stage response rate to calculate an overall ("cumulative") response rate for clients. Yet an implicit and unsupported assumption is that noncooperating and cooperating facilities contained about the same numbers of eligible clients. This information is not accessible in the DSRS reports, although it may be subject to reconstruction from the relevant public use files.

BEST COPY AVAILABLE

complete data needed to evaluate nonresponse bias at both sampling stages. This accounts for our focus on CALDATA in Section 3 of this report.

Survey differences in either first-stage or second-stage response rates might reflect either *differences in research design*—including differences in target population, sample design, follow-up procedures, duration of the follow-up interval, and other methodological differences among surveys—or *differences in degree of success in implementing the research design*. The following subsections review both the research designs and some implementation issues of the four surveys with the goal of accounting for differences in client response rates.

1. CALDATA

The sample design of CALDATA was simple relative to the designs of the other three surveys. The target population was clients who were discharged between October 1991 and September 1992 from treatment providers that received any public funding known to the State of California during that period.[7] This population included a high percentage (more than 90%) of all licensed drug and alcohol treatment providers located in California. The approach was "cold follow-up," that is, the clients were selected entirely from records obtained (as permitted by law) from treatment programs regarding specific treatment episodes. Based on information in the files, clients would then be sought, located, and recruited for interviewing about this past treatment episode and the periods before and after it.

In the first stage of sampling, information contained in the California Alcohol and Drug Data System (CADDS) was used to select a probability sample of 110 service delivery units (SDUs), selected within strata of geographic region, county, and modality of treatment. Only 106 of these units proved to have actually treated patients during the 1-year reference period, and these 106 were administered by 97 treatment provider organizations. Most of the provider organizations with more than one selected SDU offered both methadone detoxification and methadone maintenance. Moreover, several of these dual-SDU units were linked together with other sampled units owned by a few proprietary methadone "chains."

In the second stage of sampling, 87 SDUs (among 82 cooperating providers) permitted CALDATA staff to randomly select eligible clients for follow-up from their clinical records. CADDS made it possible to estimate the numbers of eligible clients in noncooperating as well as

[7] Sources of public funding included contracts with county substance abuse treatment agencies, the state Medicaid office (MediCal), or other public agencies.

in cooperating sample providers, while data collected during the record abstraction phase made it possible to compare respondents with nonrespondents within cooperating sample providers (see below).

CALDATA staff randomly selected and abstracted 3,055 records from the 87 cooperating SDUs. All clients who were discharged from treatment between October 1991 and September 1992, including those who were admitted but received no treatment services, were eligible for participation in the study. The sample also included a subsample of clients who were in methadone maintenance during the eligibility period and were still in the same episode of treatment at the time that records were abstracted in early 1993.

The 9-month interviewing field period began in April 1993 and ended in December 1993. At the conclusion of the field period, 1,858 sample clients had been interviewed; however, due to project deadlines, only 1,826 cases could be included in the published analyses and data files.[8] The postdischarge follow-up durations at the time of interview ranged from 9 to 24 months with a median of 15 months. CALDATA therefore completed interviews with 61 percent of all sampled cases (62% excluding the deceased) despite very limited identifying and locating information in the administrative records of many cooperating sample providers. The published CALDATA second-stage response rate was 60 percent (Gerstein et al., 1994).

2. SROS

Like CALDATA, SROS was a cold follow-up study that used probability sampling at both stages of sample selection. SROS was the first national-level follow-up study to employ probability sampling of providers. However, the SROS sample of cooperating providers, a total of 99 treatment facilities that had been in operation from September 1989 through August 1990, did not represent the general population of treatment providers and clients in the U.S. as comprehensively as the CALDATA sample represented California. The main reason is that SROS, fielded during a 9-month span in 1995-1996, was based on a sample of treatment facilities that had participated in the Drug Services Research Study (DSRS) in 1991. The sampling rules that had been used to select DSRS facilities from the 1990 NDATUS census of providers excluded more than half (50.4 percent) of the listed providers in NDATUS, namely those classified as treating "alcohol only" rather than "drug only" or "combined drug and

[8] The four studies did not vary in one respect: all provided the same monetary incentive of $15 for completing a follow-up interview. Three of the four studies also collected urine samples at the time of follow-up ($10 incentive) but at different sampling rates: SROS in three-fourths of all cases, NTIES in one-half, DATOS in one-quarter, and CALDATA in no cases.

alcohol" and those with missing data on this (or other) key design variables. Moreover, of the 146 facilities selected for DSRS, 47 either did not participate in DSRS and thus were ineligible for SROS (26 providers) or cooperated with DSRS but not with SROS (21 providers).

Clients in SROS were followed up 5 to 6 years after leaving treatment, compared with an average of 15 months after leaving treatment in CALDATA. As in CALDATA, the SROS identifying and locating information was restricted to information contained in abstracted clinical treatment records. The overall SROS client response rate of 59 percent, including 65 percent of living sample cases, is similar to that of CALDATA. However, whereas CALDATA absorbed about 13 field interviewer hours per completed case (including in the numerator all field interviewer hours, including those spent on noncompleted cases), SROS required close to 20 interviewer hours. Aside from the difference in resources expended, the similar results might have reflected the less urbanized character of the SROS sample and the tendency of more poorly organized programs, those with the least informative records, to be omitted from the initial sampling frame or to become lost to the sample during the intervening years. In addition, the much longer lead time of the SROS project, a result of slower stage-by-stage bureaucratic approval processes, permitted various locating efforts such as electronic search for database matches to proceed in advance of rather than relatively late in the respective 9-month field periods. Finally, since SROS was performed by the same survey organization as performed CALDATA, the experience previously gained with this cold follow-up methodology probably benefitted the second survey.

3. NTIES

The sample of substance abuse treatment programs included in NTIES was selected using purposive rather than probability sampling methods. The eligible SDUs were affiliated with one or more of 157 successful applicants to the Center for Substance Abuse Treatment (CSAT) for demonstration grants to enhance or expand treatment services for selected population groups, including individuals residing in nine of the largest urban centers ("target cities"), public housing residents, racial/ethnic minorities, pregnant and postpartum women, and adolescent and adult criminal justice populations.

Unlike CALDATA and SROS, NTIES participants were recruited to the study at the time of intake to treatment, so that follow-up was based on collecting research-oriented locator information on program records. Within the 71 cooperating and productive sample programs in 16 states, all clients were eligible for follow-up who met two minimal requirements: 1) completing a 75-minute research intake interview, which included the detailed locating

information to be used for follow-up, within 21 days after being admitted to treatment between August 1993 and October 1994; and 2) receipt of treatment services, defined as staying a minimum of one night in residential programs and completing one outpatient treatment visit beyond the intake procedure in outpatient programs. Except in the largest SDUs, where the roster of eligible clients was subsampled, all eligible clients in each sample SDU were targeted for NTIES intake interviews.

Of NTIES eligible cases, 85 percent completed the intake interview, with most of the losses due to failures to schedule the intake interview within 21 days of admission rather than to refusals. All of the 6,593 clients who completed the intake interview were targeted for follow-up interviews about 12 months after leaving treatment. In the interim, all clients were eligible for a "treatment experience" interview at the time of discharge or after an extended period of treatment, and 80 percent of the NTIES panel completed this interim interview.[9]

The 12-month follow-up response rate was 82 percent (slightly higher when the small number of deceased and other excluded cases are removed from the denominator), about 20 points higher than the follow-up interview completion rates obtained in CALDATA and SROS. Moreover, NTIES-completed cases required substantially fewer hours of follow-up interviewer time than SROS and CALDATA cases—only about 8 hours of follow-up interviewer time per completed follow-up interview. This advantage over SROS and CALDATA seems largely due to the prospective enrollment of the sample at the time of admission (involving intake interviewer effort of approximately 5 hours per case), so that the follow-up rate is based on cases for whom research-quality locator data has been collected, and who have already complied to some extent with the research protocol. The higher follow-up rate may also be partially due to the characteristics of programs included in the specialized target population; in particular, correctional programs achieved response rates exceeding 90 percent. However, the targeting of CSAT grants on "needier" programs and the concentration of sample SDUs in inner city areas would not favor the follow-up task. Moreover, if one bases the NTIES follow-up rate not on those completing the intake interview but on all those eligible for the intake interview, the follow-up completion rate is 70 percent of the eligible nondeceased sample, which is much closer to the SROS and CALDATA results.[10]

---

[9] The principal reason for noninterview was, again, missing the window of eligibility, which was within 8 weeks of discharge. Especially in outpatient programs, information about discharge was often not obtained or confirmed in time to locate and recruit the client before this window expired.

[10] The NTIES field period for follow-up interviews was approximately 12 months; however, cases were released to follow-up at different points, with some made available at 10 months after treatment with eligibility nominally ending at 14 months; others as early as 5 months after treatment due to the need to conclude the study. The

A final element in the NTIES follow-up experience was a difference in follow-up completion rates between the two survey organizations that conducted the field work. The NTIES SDUs were divided among six field assignments, four staffed and supervised by NORC and two staffed by RTI. The follow-up interview completion rate in the regions staffed by NORC and RTI were 85 percent and 70 percent, respectively.[11]

Relative to the other three major studies, NTIES under-represented methadone maintenance programs, drawing only about 8 percent of the total client sample from such programs, as compared with more than 20 percent in each of the other studies (see Exhibit II-2). NTIES was also the only one of the four studies to represent programs in correctional facilities, drawing about 23 percent of its total client sample from such facilities.

4. DATOS

Like NTIES and unlike CALDATA and SROS, DATOS featured a purposive sample of drug and alcohol treatment programs in which the follow-up research cohort was recruited on a prospective basis. In DATOS, 11 cities were initially chosen as sites for the study. Interviews were conducted only within these cities, even when sample clients had moved to other cities or nonmetropolitan areas. Within each city, an attempt was made to recruit "typical and stable" programs from each of four modalities, including short-term and long-term residential, outpatient methadone, and outpatient drug-free.

Relative to the other three surveys, an important distinguishing characteristic of DATOS is that the eligibility criteria for follow-up of clients within cooperating programs were stringent and complex, and would seem to favor higher follow-up response rates. The follow-up sample was limited to clients who 1) completed two 90-minute intake interviews and 2) were from one of the 76 programs in which 20 or more clients had completed two 90-minute DATOS intake

---

median interview took place 11 months after discharge and more than 90 percent were completed between 6 and 15 months after treatment.

[11] The difference between the completion rates of NORC and RTI might have been due to differences in difficulty of follow-up between the subsamples assigned to NORC and RTI rather than to organizational differences in follow-up effectiveness, although it is not apparent that the geographic subsamples assigned to RTI were more difficult. RTI's assignment was restricted to providers located in southern, western, and southwestern states. , NORC's assignment covered the North and Upper Midwest and included all of the older urban inner cities in the NTIES sample.

Relative to the other three surveys, an important distinguishing characteristic of DATOS is that the eligibility criteria for follow-up of clients within cooperating programs were stringent and complex, and would seem to favor higher follow-up response rates. The follow-up sample was limited to clients who 1) completed two 90-minute intake interviews and 2) were from one of the 76 programs in which 20 or more clients had completed two 90-minute DATOS intake interviews. In addition, the subsample selected for follow-up, comprised of 4,786 clients, was selected so as to oversample longer lengths of stay in treatment.

As in NTIES, some characteristics of the DATOS sample may have facilitated the locating of sample cases and, *ceteris paribus*, favored a higher follow-up response rate. These characteristics included the restriction of the follow-up to organizationally stable providers in 11 cities; intermediate research interviews for those remaining in treatment, scheduled at 1, 3, 6, and 12 months in treatment; and undersampling of clients with shorter lengths of stay, whose compliance can be more difficult to obtain.[12] As with NTIES, not all individuals admitted to treatment in the participating SDUs entered the research sample. Specific information is not available at this time on what percentage of the eligible clients completed both intake interviews.

The purposive nature of the first-stage sample and the selective noncooperation and ineligibility of sample programs reduce the generalizability or external validity of DATOS information. Simpson and Curry (1997) report that 120 cooperating treatment programs were originally selected within the 11 cities. (The number of programs selected but refusing to cooperate is not reported.) Twenty-four of these 120 programs were dropped from DATOS early on due to low initial client flow, while 20 more were excluded from the follow-up protocol because they yielded fewer than 20 clients who completed both intake interviews.

The DATOS 12-month follow-up response rate was 62 percent, about 20 percentage points lower than the comparable NTIES statistic and quite similar to the response rates for the cold follow-up in SROS and CALDATA. However, this rate falls to 48 percent of the total nondeceased participant sample, compared with 70 percent in NTIES. The DATOS protocol was less aggressive than NTIES in seeking follow-up interviews; NTIES pursued interviews within a much wider travel radius and permitted telephone interviews when personal interviews could not

---

[12] Using the criterion of remaining in treatment for 3 months or longer (very much appropriate to three of the four DATOS treatment types, less so for the short-term inpatient modality), this difference is visible in response rates among those selected for follow-up. In the long-term residential modality, 62 percent of respondents versus 50 percent of nonrespondents surpassed this length of stay; in the outpatient drug-free mode, 58 percent versus 48 percent; in methadone, 92 percent versus 78 percent.

be obtained. Telephone interviews accounted for 2 percent of NTIES follow-up cases. Nevertheless, DATOS required about 10 field interviewer hours per completed follow-up interview compared with 8 hours per follow-up interview in NTIES.

5. SUMMARY

Differences in follow-up response rates among the four treatment studies appear to be due partly to differences in research design and partly to a difference between survey organizations in follow-up effectiveness. The single interview "cold" follow-up design, operating on a purely retrospective basis, can be completed much more rapidly than the prospective/retrospective, two-interview (intake and follow-up) design. (Both DATOS and NTIES actually deployed more than two interviews.) CALDATA proceeded from sample design to comprehensive report in 18 months, while DATOS and NTIES required more than 6 years from preliminary design to publication of outcome results. (SROS was inactivated for a long period after the initial DSRS draw of 120 programs, but its subsequent active period through final report was approximately 32 months.) The trade-off for more rapid study completion is some loss in precision due to recall factors; a reduction in total information due to the reduced total interview time; and a loss of about 20 percentage points in response rate relative to a pre-enrolled panel, but only about 5 points were lost relative to the total client sample at intake. The time required for the pre-enrollment interview in the prospective/retrospective design is typically substantial. The cost of completing a post-discharge client record abstraction (required in the NTIES and DATOS protocols) is approximately equal to the cost of generating a records-only sample (CALDATA and SROS). Thus, assuming similar sample sizes and post-discharge periods, there is probably not a substantial difference in required field hours (or in associated costs) between the cold follow-up and prospective/retrospective designs.

CALDATA is the only one of the four surveys to feature a probability sample of a well-defined and geographically comprehensive general treatment population in the U.S., a population including newly established as well as long-lived and organizationally stable providers. Even though probability sampling of general treatment populations poses challenges for gaining cooperation from an adequate proportion of sampled SDUs and for successful follow-up of former treatment clients, this kind of sampling is also a *sine qua non* for rigorous comparisons of findings across studies and for cumulative development of knowledge in successive studies. Unless samples can be consistently designed to support inferences about a common population that endures in time and remains politically as well as scientifically meaningful—e.g., the population of individuals admitted to drug and alcohol treatment in a specified geographic

research results of contemporaneous surveys are merely artifacts of the different populations sampled.[13]

In the absence of a common target population, it is also hazardous to compare response rates among surveys. Since the response patterns of each survey may reflect the unique population that was represented, it is not surprising that some generalizations about differences between respondents and nonrespondents are not supported by more than one survey. For example, CALDATA reported higher response rates among Hispanics than non-Hispanics, while SROS and DATOS reported the opposite and NTIES found no differences. CALDATA, SROS, and DATOS report consistently higher follow-up rates among women than among men, whereas in NTIES the difference by gender was quite small. If all such surveys were based on probability samples of a common population, comparisons of response rates across surveys would more accurately reflect differences in measurement and follow-up methods that are not confounded with differences in target population. Knowledge of effective means of increasing the response rate would be more likely to increase with each new survey.

Even though CALDATA obtained lower total client response rates than the best results reviewed in this section (NTIES), CALDATA's application of a probability sample design to a general treatment population yielded unique information at both sampling stages, information that may be critical both in evaluating selection biases and in planning future drug and alcohol treatment follow-up surveys. The next section draws upon this advantage of CALDATA to provide the most general picture currently available of the representativeness of a large multisite treatment study.

[13] DATOS did return to largely the same cities—in many instances the same programs—that had been studied 10 years earlier in TOPS (Hubbard et al., 1989), allowing some valuable temporal comparisons of treatment components and effectiveness.

III. CALDATA: NONRESPONSE AT THE PROVIDER AND CLIENT STAGES OF SAMPLE SELECTION

Like the other recent substance abuse treatment follow-up surveys reviewed in the previous section, CALDATA employed a two-stage sampling design: first-stage selection of substance abuse treatment providers combined with second-stage selection of clients within cooperating providers. Like the other studies, CALDATA measured treatment outcomes based on retrospective reports of sample clients. Nonresponse occurred at the first stage of sampling due to noncooperation of the sample providers and at the second stage due to problems of locating and gaining cooperation from sample clients.

This section has two objectives. First, we discuss the design of CALDATA with the goal of assessing how particular design features and field operations contributed to increasing the response rate. We emphasize the important uses of administrative records obtained from sample providers and from the state of California in locating sample clients. We conclude that provider cooperation rates might be improved by developing more effective strategies to gain the cooperation of large provider chains, and client response rates might be improved by making earlier use of locating information from administrative data systems such as motor vehicle, medical eligibility, and credit bureau records.

The second objective of this section is to report the results of using administrative records to evaluate the consequences of nonresponse for the accuracy of inferences about former treatment clients in California. Our main conclusion is that nonresponse in CALDATA resulted primarily from poor-quality client-locating information obtained from providers. Nonresponse appears more highly associated with provider characteristics than with client traits that are likely to condition treatment effectiveness. Comparisons of respondents and nonrespondents using administrative records suggest few substantial differences. Yet, as discussed below, nonresponse may still bias estimates of treatment effectiveness based on CALDATA.

1. DESIGN FEATURES CONTRIBUTING TO HIGHER RESPONSE RATES

The first stage sample in CALDATA was a probability sample of California drug and alcohol treatment providers receiving funding from the State of California. A stratified random sample of 110 licensed substance abuse treatment provider units was randomly selected from a list of California-funded providers maintained by the State of California. The completed first-stage sample included 106 providers, rather than 110, because CALDATA interviewers found that 4 of the 110 providers originally sampled had no eligible clients. Providers were selected with probabilities proportional to their numbers of clients, as estimated using California

with probabilities proportional to their numbers of clients, as estimated using California administrative data, within each of five sampling strata ("modalities of treatment"), as shown in Exhibit III-1. In the second stage sampling, CALDATA interviewers randomly selected approximately 30 former clients from each cooperating provider, using a list of eligible clients developed by interviewers on-site at the facility from the administrative records of the provider. If fewer than 30 cases had been admitted during the reference year, all cases were used; if more than 30, a predetermined sampling ratio and field sampling procedures were employed; in a few very large programs, double samples (60-70) were drawn to limit variance of weights. Clients who had been discharged from the specified modality of treatment offered by the provider during fiscal year 1992 were eligible to be sampled. Interviewers then abstracted two kinds of information about the sample clients from administrative records of the provider, locating information and personal history data.

EXHIBIT III-1
NUMBER OF SAMPLE PROVIDERS BY TREATMENT MODALITY

CALDATA SAMPLE STRATUM	NUMBER OF SAMPLE PROVIDERS
1. Residential	19
2. Social model	23
3. Nonmethadone outpatient	27
4. Methadone detoxification	19
5. Methadone maintenance	18
Total CALDATA Sample Providers	106

The follow-up interview field period of CALDATA was approximately 9 months. The field period began in April 1993, about 6 months after the end of the 1-year eligibility window for discharge of eligible sample clients. The base sample comprised an estimated 3,227 individuals in the 106 selected SDUs; this number is approximate because we could not list eligible clients in noncooperating providers. This number includes all sample individuals discharged during the 1-year eligibility window and does not include the "continuing methadone sample," a supplementary sample of methadone clients that was drawn from individuals treated during the reference year but not discharged before the field period began.

Of the 3,227 sample clients, approximately 14.9 percent could not be identified for follow-up because of provider noncooperation, 18.3 percent could not be located, 9.8 percent refused to participate, and 6 percent could not be interviewed because of death, language problems, inaccessible location (even by phone), or other reasons. The number of discharge sample respondents equals 1,643.

BEST COPY AVAILABLE

The overall unweighted response rate of CALDATA—including both first-stage and second-stage nonresponse—equals 51 percent. The overall weighted response rate—calculated by multiplying each sample case by the reciprocal of its probability of selection—is lower, about 46 percent. The weighted response rate can be interpreted as the expected response rate if all individuals in the target population—rather than a sample—had been selected for follow-up.

Exhibit III-2 presents a breakdown of response rates by modality of treatment. The response rate in each modality is the product of two factors: a) the response rate based on provider cooperation, i.e., assuming all sample clients in cooperating providers were interviewed (First Stage); and b) the client response rate in cooperating providers (Second Stage). The overall response rate of 51 percent equals the product of 85 percent (the response rate based on cooperating providers) and 60 percent (the client response rate).

The First Stage of Exhibit III-2 shows that response rates based on cooperating providers were greater than 90 percent in all modalities of treatment except methadone detox (61 percent) and methadone maintenance (81 percent). The relatively low provider cooperation rate in methadone programs was due to the noncooperation by owners of two of California's large chains of private, for-profit methadone providers.

The Second Stage of Exhibit III-2 shows that the most significant factor in overall nonresponse was client nonresponse in cooperating providers. The most common source of client nonresponse was failure to locate the CALDATA sample client. Such failures had two main causes: deficient locating information obtained from providers and mobile and elusive lifestyles of some former clients. Some provider records included very incomplete or inaccurate locating information, and, in particular, most supplied too little information that could assist in locating homeless or transient clients, such as family-of-origin information or data on government program participation, including case worker name. We found that the names, addresses, and phone numbers of most relatives had limited value over a 9-month field period in establishing contact with sample clients who rarely contacted their families. Many sample clients appeared to have given fictitious names, birth dates, or Social Security numbers at the time they entered treatment, and some providers deployed few resources or interest toward validating or correcting this information. A third source of nonresponse was clients who did not wish to reopen a "closed chapter" in their lives.

EXHIBIT III-2
CALDATA — FIRST- AND SECOND-STAGE RESPONSE RATES BY MODALITY OF TREATMENT
(DISCHARGE SAMPLE ONLY)

INDICATOR	MODALITY OF TREATMENT					TOTAL
	Resident	Social Model	Nonmeth. Outp.	Meth. Detox	Meth. Maint.	
First stage. Response rate based on provider cooperation						
Sample providers	19	23	27	19	18	106
Cooperating	18	21	23	13	12	87
Sample clients	618	741	678	825	365	3227
Sample clients in cooperating SDUs	609	700	637	503	297	2746
Resp. rate based on providers	98.5%	94.5%	94.0%	61.0%	81.4%	85.1%
Second stage. Client response rate in cooperating providers						
Respondents	337	392	394	293	227	1643
Client resp. rate	55.3%	56.0%	61.9%	58.3%	76.4%	59.8%
Panel 3. Overall response rates						
Product of 1 and 2	54.5%	52.9%	58.2%	35.5%	62.2%	50.9%

BEST COPY AVAILABLE

I:\2165\08\CDNTNAL\PT1_BIAS.WPD NEDS, July 19, 1999, Page 22

To complete more than 60 percent of the cases assigned to the field (Exhibit III-2, Second Stage), CALDATA interviewers implemented a variety of creative locating approaches, including "hanging out" at homeless centers and in drug-dealing areas of urban centers. They also canvassed many kinds of administrative record systems for locating information or assistance in forwarding study "advertising"; these systems included voter registration lists, credit bureau records, jail lists, California prison locator data, vital statistic records, Veterans Administration records, death registration forms, directory assistance records, postcards and letters posted at homeless shelters and at the provider, records of contacts with shelters, motor vehicle records, and public assistance (including state Medicaid office) records.

Early in the CALDATA field period, authorization was obtained to access prison locator data from the California Department of Corrections and weekly jail lists for selected California counties. Approximately 10 percent of sample clients were found to be incarcerated. Authorization was also obtained to conduct interviews in the Federal Bureau of Prisons. Many inmates are moved frequently, and tracking them proved to be time-consuming.

2. LESSONS FROM CALDATA

Future planning of retrospective surveys of substance abuse treatment clients might benefit from the lessons of CALDATA. We think provider cooperation rates might be increased through a more strategic approach to gaining the cooperation of large proprietary provider chains. Additional steps might be to obtain authorization to access probation records as well as prison and jail lists, obtain earlier access to state motor vehicle and medical eligibility files, and carry out more frequent review of these records as they are updated during the field period.

Prospective designs, such as DATOS and NTIES, may have advantages in increasing both provider and client response rates, although differences in follow-up effectiveness may attenuate this advantage. CALDATA demonstrates that aggressively fielded retrospective treatment follow-up studies can obtain response rates that are comparable to those in successful prospective studies with follow-up periods of approximately the same duration. The main advantage of prospective studies is that, when sample clients are selected from current clients on a flow basis, locating information and pledges of cooperation can be obtained at the time clients are selected into the sample. The potential benefits of prospective surveys in increasing response must be balanced against the shorter time requirements and somewhat lower potential costs of retrospective outcome surveys. A key issue in finding the balance is the extent of bias caused by nonresponse, a topic to which we now turn.

3. NONRESPONSE BIAS

The two sources of nonresponse bias in CALDATA correspond to the two sampling stages—providers and clients.

3.1 First Stage—Bias Due to Provider Noncooperation

To evaluate this source of bias, we compared survey response distributions on a number of client and provider characteristics to corresponding distributions computed using the California subfile of the FY90-91 National Drug and Alcoholism Treatment Unit Survey (NDATUS). NDATUS also encounters provider nonresponses, so it cannot be considered a universe of which CALDATA is a subset, but a partially overlapping set of program units. However, the California subfile of NDATUS had a high estimated provider response rate relative to other states, about 95 percent in the California subfile of the FY90-91 NDATUS (Substance Abuse and Mental Health Services Administration, 1993).

Exhibit III-3 shows the results of comparisons of three client attributes, i.e., age (less than 25, 25-34, and 35 and over), sex, and ethnicity (black, non-black Hispanic, and other), and one provider characteristic, i.e., average weekly staff hours of physicians, psychiatrists, and registered nurses per 100 clients. Since CALDATA-detailed modalities cannot be precisely defined using NDATUS, each comparison in Exhibit III-3 is presented separately for two broad modalities: residential (including social model and other residential programs) and methadone (including both detox and maintenance programs). The NDATUS estimates are based on population totals for California of 423 residential and 87 methadone programs. The CALDATA estimates are weighted using selection probabilities of sample units adjusted for nonresponse, using providers as weighting cells in each stratum.

Exhibit III-3 shows that, for both residential and methadone providers, CALDATA and NDATUS distributions of clients by age, sex, and ethnicity are broadly similar. The two data sources agree that methadone clients tend to be older than residential clients, more likely to be female (especially in NDATUS), more likely to be Hispanic, and less likely to be black. The two data sources also lead to similar conclusions about the degree of staffing of physicians, psychiatrists, and registered nurses in the two kinds of programs. Both data sources estimate the level of staffing of these highly trained professionals to be approximately 6-7 times higher in methadone programs than in residential programs. These results provide little evidence that bias due to provider noncooperation is severe in the residential and methadone modalities.

EXHIBIT III-3
COMPARISONS OF CALDATA WEIGHTED SAMPLE PERCENTAGES (STANDARD ERRORS IN PARENTHESES*) WITH NDATUS

Variable	Statistic	MODALITY			
		Residential		Methadone	
		CALDATA	NDATUS	CALDATA	NDATUS
Age of clients	% < 25	13% (2)	21%	5% (1)	7%
	% 25-34	48% (3)	40%	35% (2)	32%
	% >= 35	39% (2)	39%	60% (2)	61%
Sex	% female	32% (2)	28%	37% (2)	43%
Ethnicity	% black	33% (2)	28%	9% (1)	13%
	% Hispanic	11% (2)	15%	46% (2)	37%
Weekly staff hours per 100 clients		5 (1)	6	33 (2)	44

*Standard errors are based upon an average sample design effect of 1.9– due to cluster sampling and unequal weights—and were computed using the computer program SUDAAN (Shah, Barnwell, Hunt, & LaVange, 1994).

3.2 Second Stage—Bias Due to Client Nonresponse in Cooperating Providers

The second panel of Exhibit III-2 shows that the client response rate in cooperating providers equals 62 percent or lower in every modality except methadone maintenance (76.4%). Information on detailed interview dispositions that were collected as part of the field effort indicate that the principal component of client nonresponse in every modality was failure to locate the sample client. Of 1,103 client nonresponses in cooperating providers, about 54 percent (592 nonresponses) were due to failure to locate, about 29 percent (315) were due to refusals, and about 18 percent (196) were due to death, language problems, inaccessible locations, incapacitation, and all other causes.

Exhibit III-4 presents comparisons of the characteristics of responding and nonresponding sample clients using data that were abstracted from the administrative records of cooperating providers. Panel 1 of Exhibit III-4 presents comparisons of the means of continuous variables, and Panel 2 presents comparisons of percentages. The base Ns shown in parenthesis in Exhibit III-4 refer to the numbers of CALDATA respondents and nonrespondents who had nonmissing administrative data for the variable being compared.

BEST COPY AVAILABLE

EXHIBIT III-4
CALDATA — COMPARISONS OF UNIT RESPONDENTS AND NONRESPONDENTS (BASE NS IN PARENTHESES)

STATISTIC	RESPONDENTS (N)	NONRESPONDENTS (N)
Panel 1. Means of continuous variables		
Length of stay (months)	2.8 (1,570)	2.7 (1,103)
Age at admission (years)	33.3 (1,523)	33.5 (1,068)
Education (1= < HS, 2=HS grad/CED, 3=Beyond HS)*	1.8 (1,531)	1.9 (1,090)
Number of treatment services received	2.9 (1,025)	2.8 (733)
Number of medications prescribed	1.8 (1,580)	1.9 (1,103)
Panel 2. Percentages		
% with self as primary referral source	46% (1,410)	46% (1,015)
% with legal system as primary referral source	22% (1,410)	23% (1,015)
% with public as primary payment source**	50% (1,316)	45% (871)
% female**	38% (1,585)	33% (1,103)
% black (African-American)	15% (1,578)	15% (1,103)
% Hispanic or Latino**	37% (1,319)	30% (929)
% employed at admission**	21% (1,515)	27% (1,068)
% with cocaine as primary drug**	15% (1,471)	17% (1,046)
% with heroin as primary drug**	42% (1,471)	40% (1,046)
% with alcohol as primary drug**	27% (1,471)	29% (1,046)
% completing treatment plan**	32% (1,643)	31% (1,103)
% with aftercare plan stated in record	35% (1,643)	35% (1,103)

* Significant difference based on two-sample t test, two tail, $\alpha = .05$.

**Significant difference based on chi-square test, $\alpha = .05$.

BEST COPY AVAILABLE

The main conclusion from Exhibit III-4 is that few measured variables evidence substantial differences between respondents and nonrespondents. Even statistically significant differences, as gauged by two-sample t-tests for comparisons of continuous variables (Panel 1) and chi-square tests for comparisons of percentages (Panel 2), tend to be substantively small. The large sample sizes portend that even small differences will be significant at conventional levels. Two of the largest (though still relatively modest) differences in Exhibit III-4, primary payment source (50% public vs. 45%) and Hispanic ethnicity (37% vs. 30%), are based on program variables with item nonresponse rates greater than 20 percent. The program data are more complete, however, for gender (38% female vs. 33%) and employment at admission (21% vs. 27%). Women typically respond to surveys at a higher rate than men, which holds in this population as in others. The lower response rates of privately paying, employed, and white non-Hispanic sample persons are somewhat surprising. Comments received from some refusers, such as the comment that substance use and treatment comprised a "closed chapter" that they did not choose to revisit in an interview, suggest the possibility of deliberate concealment. If this interpretation is correct, there would be a mild bias toward exclusion of relatively higher income individuals who, by and large, would be expected to have better treatment prognoses.

In summary, analysis of CALDATA nonresponse at the two stages of sampling produced evidence of only modest potential biases based on measured characteristics. Comparisons of CALDATA to NDATUS (Exhibit III-3) and of CALDATA respondents and nonrespondents (Exhibit III-4) suggest that respondents and nonrespondents are similar in demographic characteristics. Exhibit III-4 is especially compelling because of the variety of characteristics that were measured, including measures of treatment services and pre-treatment and within-treatment substance use.

Two hypotheses to account for the small differences between respondents and nonrespondents are as follows:

a) In an aggressively fielded follow-up study, nonresponse at the level of individual clients results primarily from poor-quality addresses and other locating information (criminal justice, hospital, Social Security, etc.) and secondarily from differential nonresponse by higher income individuals;

b) The quality of locating information may be largely independent of social attributes of clients, with the exception noted above. This suggests nonresponse might be largely independent of treatment outcomes. However, as discussed in section 5, caution is warranted.

IV. Bias Due to Selection into Treatment

Individuals who enter treatment in the U.S. are not a random sample of the general population, of the population using substances, or of the population who have substance abuse disorders according to standard diagnostic criteria. Between one-half and two-thirds of individuals entering treatment for a substance abuse problem in the U.S. enter treatment at least in part due to pressure from the criminal justice system (Hubbard et al., 1989; Pringle, 1982; Schildhaus et al., 1998), while the remainder enter treatment of their own volition ("self-select") or because of pressures from other sources.

Carroll and Rounsaville (1992) compared treated cocaine abusers who met Research Diagnostic Criteria (RDC) for cocaine dependence with matched untreated cocaine abusers, and found that, on average, the untreated individuals had higher levels of polysubstance use, fewer social supports (such as familial and employment ties), fewer familial and employment problems resulting from cocaine abuse, and greater involvement in illegal activities. Rounsaville and Kleber (1985) employed a similar research design to compare treated and matched untreated opiate addicts and reached similar conclusions, except that—unlike untreated cocaine abusers—untreated opiate addicts had rates of psychiatric disorder lower than those of their treated counterparts and levels of illegal activity that were no higher.

Neither Carroll and Rounsaville (1992) nor Rounsaville and Kleber (1985) found evidence that entering treatment is primarily a function of the level of substance use. Treated and untreated cocaine abusers reported similar levels of cocaine use (Carroll and Rounsaville, 1992), and treated and untreated opiate addicts reported similar levels of opiate use (Rounsaville and Kleber, 1985).[14] Similarly, Hser, Maglione, Polinsky, & Anglin (1998) compared treated with untreated individuals who had been referred to treatment and found no significant differences in type of drug use or years of use. Like Carroll and Rounsaville (1992) and Rounsaville and Kleber (1985), Hser et al. found that untreated individuals tended to have fewer familial and economic problems. Hser et al. also concluded that, on average, untreated individuals had lower levels of psychological distress.

---

[14] These results contradict findings of previous studies that did not match treated with untreated individuals according to RDC or comparable criteria for substance dependence (Chitwood and Morningstar, 1985; Graeven and Graeven, 1983; Price, Cottler, & Pearl, 1990).

Most published comparisons of treated and untreated substance abusers, including the studies mentioned in the preceding paragraphs, have been limited to local area and convenience samples, usually obtained by matching individuals in treatment at one or more facilities to untreated individuals identified by the individuals who are in treatment. Thus, most previous studies cannot be used to make inferences about the selection process affecting treatment entry in the U.S. population as a whole.

Two possible exceptions are Schutz, Rapiti, Vlahov, & Anthony (1994) and Gerstein, Foote, & Ghadialy (1997). Schutz et al. used community outreach techniques to recruit injecting drug users (IDUs) in Baltimore who had not been in treatment for at least 1 year and followed the recruited IDUs over time to observe subsequent patterns of entry and nonentry into treatment. Schutz et al. concluded that recent drug overdose, relatively high frequency of injecting drugs, and prior treatment or arrest history predicted entry into detoxification. Living with a spouse or other partner, being female, long duration of drug use, and prior treatment predicted entry into methadone maintenance. Given that selection may depend on sociocultural context and on characteristics of treatment services and policies, in particular metropolitan areas (e.g., Hartnoll, 1992), drawing national conclusions based from Schutz et al. may be problematic.

Gerstein, Foote, & Ghadialy (1997) used data from the 1992-93 National Household Surveys on Drug Abuse (NHSDA)—a national probability sample of the U.S. noninstitutionalized population aged 12 and older—to compare NHSDA respondents who reported ever receiving substance abuse treatment with other NHSDA respondents. Only 0.7 percent of the NHSDA surveyed population, a total of about 1.4 million individuals, had received treatment for a drug problem in the past 12 months, and 2.3 percent had ever received treatment. Relative to the NHSDA surveyed population, the population ever receiving treatment was composed largely of individuals who reported early initiation of alcohol and marijuana use and high levels of recent drug use. Gerstein, Foote, & Ghadialy (1997) also reported discrepancies in limited comparisons that could be made—using NHSDA measures of age, gender, and race/ethnicity—between the demographic profiles of NHSDA respondents currently in treatment and those in provider-based surveys, including NDATUS[15] and DSRS. The main discrepancy pertained to the percentage of treatment clients who were Hispanic, about 8 percent in NHSDA as compared with 12 percent in NDATUS and DSRS.

[15] Since 1995, NDATUS has been called the Uniform Facility Data Set (UFDS).

By combining data from sufficient numbers of successive NHSDAs, preferably all NHSDAs conducted since 1992 (the only NHSDAs including questions on substance abuse treatment), it might be possible to obtain a large enough sample size to compare treated substance users—a rare group in the surveyed population—and nontreated individuals with comparable levels of substance dependence—another rare group. Such a comparison might provide a basis for more accurate conclusions about the factors affecting selection into treatment in the U.S. household and general populations.[16]

Exhibit IV-1 illustrates the potential bias in estimates of treatment effectiveness that is due to nonrandom selection into treatment.[17] Exhibit IV-1 is a scatter plot for a response variable Y and an explanatory variable X, where the dots represent sample observations under the assumption of simple random sampling with no selection bias. For example, Y might be a measure of treatment effectiveness constructed by differencing post-treatment and pre-treatment indicators of substance use, criminal activity, or income and employment. X might be a measure of pre-treatment psychological health or well-being. The example is realistic because psychological distress has been identified both as a factor predisposing individuals to enter treatment (e.g., Hser et al., 1998) and as a factor associated with relatively poor treatment outcomes (e.g., Gerstein, Datta et al., 1997). We assume that both X and Y are measured on a scale ranging from 0 to 100 and that missingness is high when X is less than about 40.

The solid line in Exhibit IV-1 (before adjustment) represents the estimated slope of Y on X in the absence of bias due to nonrandom selection into treatment. The cross-hatched area in the lower lefthand corner of the figure shows the area in which missingness is high. The dotted line (after adjustment) shows the estimated slope of Y on X after taking the bias due to nonrandom selection into account. The new regression line implies that the original (true) regression no longer fits the data. In the dotted line, adjusted for nonrandom selection, the estimated slope of Y on X is seriously underestimated. Depending on the missing data

---

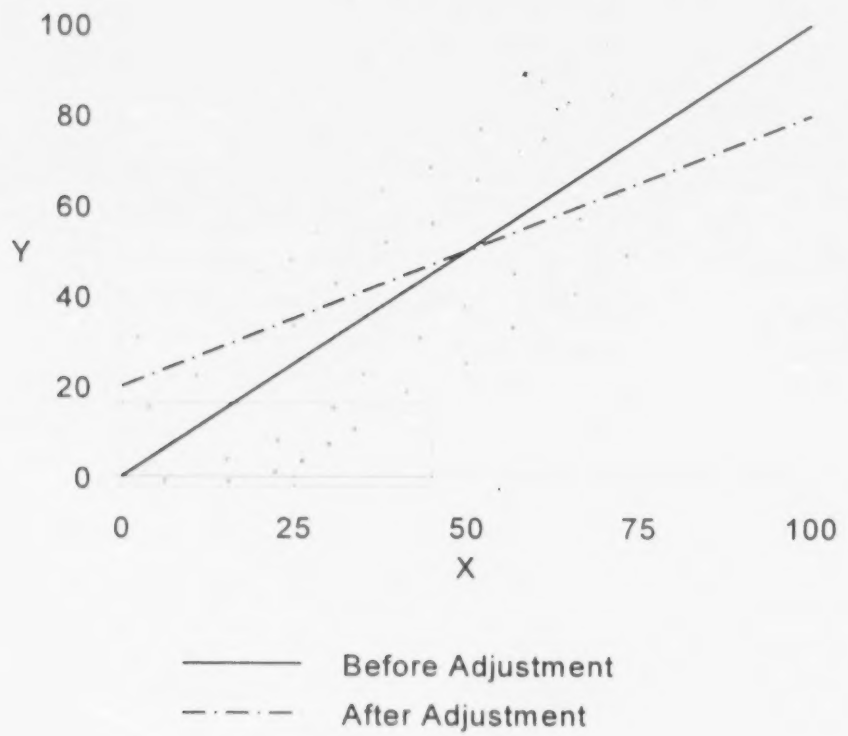
[16] As a household survey, the NHSDA represents about 98 percent of the population aged 12 and older in the U.S. SAMHSA, 1997b, Chap. 1). However, the NHSDA does not represent active military personnel, individuals living in institutional quarters (e.g., prisons, nursing homes, treatment centers), and those with no permanent residence (e.g., homeless people). Thus, the NHSDA may be inappropriate for general population inferences to the extent that the process of selection into treatment differs between the household and non-household subpopulations and to the extent that the non-household subpopulation accounts for a substantial fraction of total individuals in need of treatment. These questions would need to be addressed as part of the proposed application of NHSDA.

[17] Exhibit IV-1 is adapted from Berk (1983).

BEST COPY AVAILABLE

EXHIBIT IV-1

BIAS DUE TO SELECTION INTO TREATMENT



mechanism, e.g., on whether low or high values of X have high missingness, the bias might be positive rather than negative.

External validity has been undermined, and this consequence would perhaps not be too surprising to many treatment researchers. It is unlikely that many researchers would use data from a sample of treatment clients to draw conclusions about persons in need of treatment in the general population or about individuals with levels of substance abuse severity that are comparable to those of individuals in the treatment population.

What is less commonly recognized is that *internal validity is also compromised.* Even if the findings explicitly state that results apply only to persons who enter treatment, or that the results apply only to persons with high values of X, the findings may still be faulty. This unwelcome conclusion follows from Exhibit IV-1 together with one of the key assumptions of

From Exhibit IV-1, it is apparent that, while the residuals are approximately uncorrelated with corresponding values of X in the case of the solid line, this is far from being true in the case of the dotted line. Rather than being uncorrelated, the "after" residuals—equal to the vertical distances between data points and the dotted line—are highly positively correlated with X. The higher the value of X, the higher the average value of the residuals calculated at that value of X. This pattern of residuals violates the key assumption of regression, and it implies that the estimated effect of X on Y will be biased, *even if an analyst is only interested in the effect of X on Y in the population of individuals who enter treatment.* An intuitive explanation is that, given the correlation between residuals and X, causal effects are attributed to X that actually resulted from unmeasured or omitted factors, so that the effect of X on Y might be seriously overestimated, even if the concern is strictly with the population entering treatment.

In summary, the selection bias problem cannot be dismissed by restricting the scope of conclusions to the nonrandom subset of potential treatment beneficiaries who enter treatment or to the subset of such clients with high or low values on a specified variable.

V. METHODS FOR ESTIMATING AND EVALUATING SELECTION BIAS

The preceding sections have reviewed two selection processes or "missing data mechanisms" that potentially give rise to bias in estimates of treatment effectiveness from substance abuse treatment follow-up studies. They are the process of entry into treatment and the process of survey nonresponse among potential sample providers and potential sample clients. Following theories of missing data in statistics and econometrics, it makes sense to regard both selection into the treatment population and survey nonresponse as processes that can truncate, or otherwise distort, the observed distributions of variables. "Missingness," the probability that a response is missing for a variable, can depend on unmeasured as well as measured characteristics.

Some reports on treatment follow-up surveys include comparisons of the distributions of respondents and nonrespondents on variables that are measured for both, such as the comparison shown in Exhibit III-4. While informative, these kinds of comparisons cannot fully establish the absence of selection bias even if few or no differences between respondents and nonrespondents are detected. On the contrary, the distinction between biases that can and cannot be detected by controlling for measured characteristics is central to research on missing data.

Missing data are said to be "missing at random" (MAR) if any important differences between respondents and nonrespondents, or between individuals who enter and do not enter treatment, can be captured by variables that are measured for both. (See Heckman, 1976, 1979; Rubin, 1977, 1987; Berk, 1983; Maddala, 1983; Little and Rubin, 1987; Heckman and Hotz, 1989; Winship and Mare, 1992; Little and Schenker, 1995; and Stolzenberg and Relles, 1997.) For example, suppose that the outcome or response variable in an analysis of treatment effects is the change in employment status between the pre-treatment and post-treatment reference periods. If nonresponse depends only on treatment modality, and if modality is measured for both nonrespondents and respondents, then selection bias can be controlled by an analysis that stratifies on modality or adjusts for modality in some other fashion, such as a weighted analysis using weights adjusted for percentages responding in different modalities. A term often used as a synonym for "MAR" is "ignorable." If the missing data are MAR, then the missing data mechanism can safely be ignored in making inferences from the data, provided that the analysis controls for measured characteristics associated with the response rate.[18]

---

[18] A special situation when MAR does not imply "ignorable" is when the parameters of the missing data mechanism are the same as—or mathematically related to—the parameters of the substantive process that determines the outcome variable. See Rubin (1976) for details.

If the missing data are not MAR, then the missing data mechanism is "nonignorable." Continuing the example of the previous paragraph, one possibility is that nonresponse on the change-in-employment-status outcome variable depends not only on type of treatment but also on the outcome variable itself. For example, this would be true if, among persons needing treatment, employed persons were more likely to enter treatment than unemployed persons or persons not in the labor force. In this situation, the missing data are *not* MAR because missingness depends on a variable that is itself sometimes missing. Ignoring the missing data mechanism—by analyzing responding cases as if they comprised a stratified random sample of the survey population—could result in biased estimates of treatment effects. Unbiased estimation generally requires developing an explicit model of the missing data mechanism.

In substance abuse treatment follow-up studies, there are often grounds for suspecting that the MAR assumption is violated with respect to both entry into treatment and nonresponse. For example, individuals whose substance use behaviors give rise to familial or employment problems may be more likely to enter treatment than other potential clients (Carroll and Rounsaville, 1992), and measures of such problems are typically unavailable for persons who do not enter treatment. Similarly, sample clients who experience less successful treatment outcomes may be less likely to be located and to respond to follow-up interviews than those with more successful outcomes. Such differences might be present within each subclass of clients that can be defined using measured covariates. If so, there exists a bias that cannot be discerned by comparing respondents and nonrespondents. Estimated treatment effects might be biased even if all measured covariates were controlled in an analysis.

Despite the potential for bias, published analyses of substance abuse treatment follow-up surveys typically make an assumption that is even stronger than MAR. The usual analytical approach is "complete-case analysis," also known as "listwise deletion," which means that cases with incomplete data, including both cases without responses on one or more follow-up variables and cases with no follow-up interviews, are discarded for the purpose of the analysis. Complete case analysis yields unbiased estimates of treatment effects only if the missing data mechanism is "missing completely at random" ("MCAR"), i.e., uncorrelated with all measured as well as unmeasured variables. Complete-case analysis assumes that *all* missing data—including data missing due to nonrandom entry into treatment as well as to survey nonresponse—are MCAR. Given the potential selection biases discussed in preceding sections, the MCAR assumption seems unlikely to be satisfied in many applications.

Future analyses of substance abuse treatment follow-up surveys should apply statistical methods that make more realistic assumptions about the selection processes giving rise to the

observed data, methods that have developed rapidly since the 1970s. Good summaries are Little and Rubin (1987) and Little and Schenker (1995). Three statistical approaches that merit consideration are discussed in the following subsections—weighting adjustments including poststratification, likelihood-based estimation, and imputation.

1. WEIGHTING ADJUSTMENTS INCLUDING POSTSTRATIFICATION

A simple modification of complete-case analysis is to assign a selection weight to respondents to reduce or eliminate biases due to nonrandom entry into treatment and due to nonresponse. Weighting for nonresponse is common in many Federal surveys and is carried out in two stages: First, adjustment cells are formed based on background characteristics measured for both respondents and nonrespondents. Second, the weight assigned to a particular respondent equals the inverse of the response rate in the adjustment cell containing the respondent.

While weighting for nonresponse is simple and sometimes efficacious in reducing bias, it is only effective if nonresponse varies significantly according to variables that are measured for both respondents and nonrespondents, i.e., if the missing data can be assumed to be MAR. Moreover, this approach does nothing to reduce bias due to nonrandom entry into treatment, because data from substance abuse treatment follow-up surveys are only available for individuals who were admitted to—and in some surveys, discharged from—treatment.

A more promising approach for substance abuse treatment follow-up surveys might be a weighting approach called *poststratification*, which weights respondents to match the distribution of variables available from an external data source. The poststratification variables do not need to be known for nonrespondents, and the external data source can represent population in need of treatment rather than population admitted to or discharged from treatment. Although poststratification to client distributions in the Uniform Facility Data Set (formerly NDATUS) has long been possible, application of poststratification to the population in need of treatment has until recently been impossible because of the absence of external data on this population in the U.S. As discussed in Section 2, the combined samples of the National Household Surveys on Drug Abuse (NHSDA) conducted since 1992 might provide sufficient external data to construct poststratification weights for the U.S. household population in need of treatment.

2. LIKELIHOOD-BASED METHODS

Likelihood-based methods for statistical analysis with missing data are extensions of the familiar "maximum-likelihood" method of statistical estimation that is discussed in introductory

textbooks in statistics. In multiple regression analysis, for example, maximum likelihood yields estimates that are equivalent to the more familiar "ordinary least squares" (OLS), provided that standard assumptions—primarily, the assumption of independent and identically distributed errors and the assumption of zero correlation between error and explanatory variables—are satisfied and provided also that the response variable is normally distributed.

The likelihood-based approach to estimation with missing data comprises two principal bodies of techniques: methods assuming an ignorable (MAR) missing data mechanism and methods assuming a nonignorable missing data mechanism. Ignorable methods assume that missingness depends only on variables that are observed for both respondents and nonrespondents. Nonignorable methods assume that missingness also depends on variables that are unmeasured for nonrespondents or for both respondents and nonrespondents. Unlike ignorable methods, nonignorable methods require the formulation and estimation of an auxiliary statistical model for each postulated missing data mechanism.

Even though there are grounds for suspecting that missing data mechanisms in treatment follow-up studies are non-MAR and nonignorable, ignorable likelihood-based have two important advantages at the current stage of research: First, specifying appropriate models for missing data mechanisms is difficult, and nonignorable methods with incorrectly specified missing data mechanisms can yield results that are far inferior to those of ignorable methods. Second, realistic nonignorable models tend to be complex. Even if such models are correctly specified, available data are often insufficient to yield accurate estimates of model parameters. Little and Schenker (1995) provide a more detailed discussion of advantages and disadvantages of alternative likelihood-based methods. Even though ignorable models make assumptions that may be dubious in applications to substance abuse treatment follow-up surveys, these models are still more realistic than complete-case analysis.

Although ignorable methods are generally preferable, there is a famous nonignorable model that merits consideration in substance abuse treatment follow-up studies. This is the ingenious "probit selection model," also called the "random censoring model," of Heckman (1976). Applying this model to nonresponse bias in treatment research requires that two equations are correctly specified:[19]

---

[19] In addition, Heckman's model assumes that Y_1 has a normal distribution with constant variance and that Y_2 has a Bernoulli distribution. These are important assumptions. It may be possible to transform the outcome variable to better approximate the normality assumption.

a) The "treatment outcome equation"—the linear regression of a continuous treatment outcome measure "Y_1"—such as before/after reduction in monthly substance use—on explanatory variables that may be characteristics of the client, provider, and/or treatment services received

b) The "selection equation"—the probit regression[20] of a binary (0-1) outcome variable "Y_2"—whether or not data are missing for the outcome variable—on explanatory variables that may be characteristics of the client, provider, and/or treatment services received.

For illustration, we assume that each equation has a single explanatory variable:

treatment outcome equation: $Y_1 = a_0 + a_1 X + u$

selection equation: $Y_2 = F(b Z + v)$,

where u and v are independent errors and F denotes the cumulative normal distribution.[21]

Even in this simplified model, we can choose the single explanatory variable in each equation based on previous research: Suppose that the object is to correct for bias due to client nonresponse, let X denote the duration of treatment in months, and let Z denote a quantitative measure of the quality of locating information that is available from the provider. Research suggests that the parameter a_1 in the treatment outcome equation—the effect of duration on outcome—should be positive and that the parameter b in the selection equation—the effect of locating information on the probability of nonresponse—should be negative. However, estimates of a_1 that are obtained without taking into account the selection equation might be badly biased.

Using the two equations together, Heckman (1976) shows how to obtain a consistent (large-sample unbiased) estimate of a_1 by means of standard probit regression and least-squares estimation procedures. (Details of the estimation are also presented, along with reviews of related research, in Maddala, 1983; Little and Rubin, 1987; and Stolzenberg and Relles, 1997.) Subsequent evaluations of Heckman's estimation procedure based on simulations suggest that the method can be unstable and sometimes yields contradictory results, such as negative predictions for outcome variables known to be positive (Stolzenberg and Relles, 1990).

---

[20] The probit regression is similar to logistic regression, except that the cumulative normal distribution (inverse probit), rather than the cumulative logistic, is used to scale predictions based on the model.

[21] For simplicity, the treatment outcome equation also assumes that the outcome variable has been scaled to have unit variance.

BEST COPY AVAILABLE

Stolzenberg and Relles (1997) present a programmatic approach for assessing the utility of Heckman's correction for selection bias in specific applications.

For the illustrative outcome and selection equations presented above, the key issue in determining the applicability of Heckman's model is the magnitude of the correlation between X and Z. If the correlation between treatment duration (X) and locating information (Z) is *moderate* in magnitude (say, in the range between 0.3 and 0.6), then Heckman's estimation procedure is likely to improve the estimation of a_1. On the other hand, if this correlation is either too low or too high, Heckman's procedure will either have little effect or will worsen rather than improve the estimation. Previous research suggests that X and Z are probably correlated positively, and the magnitude of this correlation seems likely to range between moderate and strong. Data from CALDATA and other treatment follow-up surveys might be reanalyzed to assess the utility of Heckman's correction for bias due to nonresponse.

Heckman's approach is also potentially applicable to the problem of bias due to nonrandom selection into treatment. Such an application would require data on the population in need of treatment. The pooled 1992-1997 NHSDAs might be a good source of the kinds of data that are needed to assess this application.

3. IMPUTATION

Another statistical approach that has experienced rapid development in recent years involves imputing a value for each missing data value. The key advantage of imputation is restoration of the rectangular form of the data matrix, so standard methods of statistical analysis can be applied to the completed data set. The imputation procedure can also be carried out once and for all—preferably by the data producer—so that subsequent secondary analyses can use a common completed data set.

The principal line of advance has been from traditional deterministic methods of imputation to random or stochastic methods. For example, given missing data on Y, a *deterministic* regression imputation first uses completed cases to estimate the regression of Y on a battery of predictors—X_1, X_2, ..., X_k—and then assigns predicted values based on the estimated regression equation to cases with missing values of Y. A *random* regression imputation uses a similar model, except that each missing value is replaced by its regression prediction with a random error added on, and the random error has variance equal to the estimated residual variance around the regression hyperplane. Unlike the deterministic procedure, the random procedure preserves the original variability of imputed variables in the completed data set, as

opposed to regressing each imputed value toward its predicted conditional mean based on the imputation model.[22]

An important principle of random imputation is to use as many correlated observed predictors as computationally feasible in carrying out the imputations (Little, 1988). The use of multiple levels of analysis—including the treatment episode, client, and provider levels in treatment follow-up surveys—is also recommended, because covariates operating at each level can be predictive of client outcomes with missing values. Multilevel (hierarchical) models can be used to realistically reflect the hierarchical structure—sample clients nested within sample providers—of substance abuse treatment follow-up data. Using multivariate multilevel models, one can simultaneously and randomly impute a vector of two or more outcome variables with missing values, using all available variables and levels of analysis in each imputation equation. Such models have already been applied in imputing missing data values in surveys of students nested within schools (Goldstein, 1995).

In analyzing treatment follow-up surveys, the imputation strategy using multivariate multilevel models can work not just for cases with scattered item nonresponses. Given client-level predictors collected in the baseline interview, provider-level predictors, and treatment episode-level predictors, the strategy can also yield improved results for cases that were unit nonresponses in the follow-up interview. Future research might compare results on treatment effectiveness already reported for CALDATA or other substance abuse treatment follow-up surveys—obtained using traditional complete-case analysis—with the results of the similar analyses applied to data sets that were first filled-in using multivariate multilevel models.

---

[22] An important extension of random imputation is "multiple imputation" (Rubin, 1987), which produces multiple random imputations—based on different random draws from the stochastic error distribution of the model—for each missing value. The advantage of multiple imputation is the realistic assessment of imprecision in statistical inferences that is due to the imputation procedure itself. The imputation model that is used in generating random imputations must be general enough to include all of the models that are of interest in the substantive research as special cases.

VI. Implications for Treatment Research, Policy, and Practice

Analysis of the four major multisite substance abuse treatment follow-up studies completed to date in the 1990s indicates that nonrandom treatment entry, nonrandom sampling of providers, and provider and client nonresponse represent important challenges to the validity and generalizability of study findings. The problem caused by nonrandom sampling of providers is ameliorated when probability sampling is used to select providers, as in SROS and CALDATA. Detailed comparisons of findings across the follow-up studies, and comparisons of the follow-up studies with NDATUS (now the Uniform Facility Data Set), can help in assessing the seriousness of selection biases due to nonrandom entry into treatment and nonresponse.

The resources that have been devoted to reducing provider and client nonresponse in the four studies reported here may or may not represent the limits of what one can practically expect to find available for large-scale studies, although smaller methodological efforts to study the cost-effectiveness of larger or smaller efforts, such as incentive differences or substantially reduced or elongated field periods, would be useful. The present, demonstrated best practice in follow-up response rates in such large-scale substance abuse treatment studies is in the neighborhood of 85 percent of providers in a randomly selected provider cohort; 65-70 percent of the total (nondeceased) admission cohort—equivalent to 80-85 percent of an intake-inducted panel—in cooperating providers; and thus about 55-60 percent of the total admission cohort in a full probability sample of provider-distributed clients.

The experience of CALDATA suggests that noncooperation of large multi-site proprietary chains is a significant potential limitation to provider response rates and thus to the generality of research findings. Some correlates of client nonresponse and of treatment entry, with implications for the findings of treatment outcome studies, have been identified in the multi-site as well as smaller scale studies. For example, participation in follow-up by Hispanics seems to be volatile; they were more compliant with follow-up than non-Hispanics in CALDATA, less compliant in SROS and DATOS. Several studies indicate that fully employed persons are both more likely to enter treatment and less likely to comply with follow-up protocols.

Several lines of research using advanced statistical methods might be explored to assess and correct biases due to nonrandom entry into treatment and to nonresponse:

- Use combined samples of the National Household Survey on Drug Abuse (NHSDA) conducted since 1992 to compare treatment clients with other chronic drug users

- Use combined samples of the NHSDA conducted since 1992 to poststratify estimates of treatment effectiveness from substance abuse treatment follow-up studies
- Assess the utility of Heckman's (1976) correction for nonresponse bias by applying the programmatic methods of Stolzenberg and Relles (1997) to CALDATA
- Assess the utility of Heckman's (1976) correction for bias due to nonrandom entry into treatment by applying the programmatic methods of Stolzenberg and Relles (1997) to data from the pooled 1992-1997 NHSDA
- Revise key CALDATA outcome analyses using the completed data set based on a multivariate multilevel imputation model, and compare the revised results with the original ones.

The value of research results to policy makers and to the general public can be no greater than the quality of the data upon which the results are based. Public resource allocations to drug treatment in contrast to other instruments of drug control policy, and the priority of drug control in general, are responsive to policy studies on treatment effectiveness, cost-effectiveness, and cost-benefits (cf. Caulkins et al., 1999; cf. Manski, Pepper, & Thomas, 1999). All such studies make extensive use of survey results to calibrate and anchor their models. Survey data are known to be affected by many sources of error, including sampling errors, measurement errors, processing errors, and erroneous assumptions in the statistical models that are used to summarize the data, as well as errors due to nonresponse and selection. Each error source affects the quality of the data, and consequently the probity of conclusions that are based upon the data. The research literatures of statistics and other fields are replete with examples of how neglect of one or more sources of survey error can give rise to faulty conclusions (e.g., Groves, 1989), and it behooves researchers to minimize such errors by purging them where possible and adjusting for them as appropriate.

Evaluations of data quality help to identify the important sources of error in a particular kind of survey and suggest strategies for reducing the error in future studies. Data quality evaluations also alert the consumers of research products that there are potential problems with the products, just as warnings affixed to other products by authority of Federal agencies such as the Food and Drug Administration help to alert consumers. Nonresponse is an important potential source of bias in treatment follow-up studies, because the overall response rate in these studies is probably no greater than 60 percent, which is lower than is obtained in many other kinds of surveys (Groves, 1989), although much higher than is often obtained in highly regarded political polling and market research. Selection into treatment is another important potential source of bias, because it is creates uncertainty as to whether findings based on treatment follow-

up studies can be applied with confidence to new treatment cohorts or to prospective client populations who might benefit from treatment services.

Treatment practice is conservative and tends to change slowly in response to outcome research. Nevertheless, large-scale outcome studies affect clinical management practices such as pretreatment medical examinations, the use of case managers, and staging of treatment. These studies also contribute to pressures on inpatient utilization and prescription of brief courses of treatment, and they lend empirical strength—or weakness, as the case may be—to initiatives to provide specialized or matched services to particular population groups that are often considered to be underserved or less successful in treatment, including groups defined by demographic features, primary substance, or the presence of comorbid conditions. Recognizing and reducing nonresponse error and selection bias in large-scale outcome studies will improve the accuracy of findings and help assure that changes in clinical practice will not simply reflect trends in the managed care marketplace or the political arena but will also make clinical work more effective.

REFERENCES

Andersen, R. (1995). Revisiting the behavioral model and access to medical care: Does it matter? Journal of Health and Social Behavior, 36, 1-10.

Andersen, R., & Newman, J. (1973). Societal and individual determinants of medical care utilization in the United States. Milbank Memorial Fund Quarterly Journal, 51, 95-124.

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. American Sociological Review, 48, 386-398.

Carroll, K., & Rounsaville, B. (1992). Contrast of treatment-seeking and untreated cocaine abusers. Archives of General Psychiatry, 49, 464-471.

Caulkins, J.P., Rydell, C.P., Everingham, S.E., Chiesa, J., & Bushway, S. (1999). An ounce of prevention, a pound of uncertainty. Santa Monica, CA: Rand.

Chitwood, D., & Morningstar, P. (1985). Factors which differentiate cocaine users in treatment from nontreatment users. International Journal of the Addictions, 20, 449-59.

Cochran, W. (1972). Sampling techniques. Second edition. New York, NY: Wiley.

Cunningham, J., Sobell, L., Agrawal, S., & Toneatto, T. (1993). Alcohol and drug users' reasons for seeking treatment. Addictive Behaviors, 18, 347-353.

Gerstein, D. R., & Harwood, H. J. (Eds.) (1990). Treating drug problems: A study of the evolution, effectiveness, and financing of private and public drug treatment systems. Washington, D.C.: National Academy Press.

Gerstein, D. R., Johnson, R. A., Harwood, H. J., Fountain, D., Suter, N., & Malloy, K. (1994). Evaluating recovery services: The California drug and alcohol treatment assessment (CALDATA). Sacramento, CA: Department of Alcohol and Drug Programs.

Gerstein, D., Foote, M., & Ghadialy, R. (1997). The prevalence and correlates of treatment for drug problems. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies, April 1997.

Gerstein, D. R., Datta, A. R., Ingels, J. S., Johnson, R.A., & Rasinski, K.A. (1997). Final Report: National Treatment Improvement Evaluation Survey. Chicago, IL: National Opinion Research Center.

Goldstein, H. (1995). Multilevel statistical models. London: Edward Arnold.

Graeven, D., & Graeven, K. (1983). Treated and untreated addicts: factors associated with participation in treatment and cessation of heroin. Journal of Drug Issues, 13, 207-18.

Groves, R., (1989). Survey errors and survey costs. New York, NY: Wiley.

Hartnoll, R. (1992). Research and the help-seeking process. British Journal of Addiction, 87, 429-437.

Heckman, J. (1974). Shadow prices, market wages and labor supply. Econometrica, 42, 679-94.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement, 5, 475-492.

Heckman, J. (1979). Sample selection bias as a specification error. Econometrica, 45, 153-161.

Heckman, J., & Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. Journal of the American Statistical Association, 84, 1078-1103.

Hubbard, R., Marsden, M., Rachel, J., Harwood, H., Cavanaugh, E., & Ginsburg, H. (1989). Drug abuse treatment: A national study of effectiveness. Chapel Hill, NC: University of North Carolina.

Hser, Y., Maglione, M., Polinsky, M. & Anglin, M. D. (1998). Predicting drug treatment entry among treatment-seeking individuals. Journal of Substance Abuse Treatment, 15, 213-220.

Johnson, R. A., Gerstein, D., & Rasinski, K. (1998). Adjusting survey estimates for response bias. Public Opinion Quarterly, 62, 354-347.

Kish, L. (1965). Survey sampling. New York, NY: Wiley.

Little, R., & Rubin, D. (1987). Statistical analysis with missing data. New York, NY: Wiley.

Little, R. (1988). Missing data adjustments in large surveys. Journal of Business and Economic Statistics, 6, 287-301.

Little, R., & Schenker, N. (1995). Missing data. In G. Arminger, C. Clogg, & M. Sobel (Eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 39-76). New York, NY: Plenum.

Maddala, G. S. (1983). Limited dependent and qualitative variables in econometrics. Cambridge, U.K.: Cambridge University Press.

Manski, C.F., Pepper, J.V., & Thomas, Y.F. (Eds.) (1999). Assessment of two cost-effectiveness studies of cocaine policy. Washington, D.C.: National Academy Press.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). Applied linear statistical models, fourth edition. Chicago, IL: Irwin.

Office of National Drug Control Policy (1998). The national drug control strategy. Washington, D.C.: Executive Office of the President.

Price, R., Cottler, L., & Pearl, R. (1990). Drug abuse treatment utilization in a general population: Prevalence, needs assessment, and predictors. Committee on Drug Dependence, 52nd annual meeting, Richmond, VA.

Pringle, G. (1982). Impact of the criminal justice system on substance abusers seeking professional help. Journal of Drug Issues, 12, 275-283.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. Marsden (Ed.) Sociological methodology 1995 (pp. 111-164). Washington D.C.: Am. Soc. Association.

Rounsaville, B., & Kleber, H. (1985). Untreated opiate addicts: How do they differ from those seeking treatment? Archives of General Psychiatry, 42, 1072-1077.

Roy, A. (1951). Thoughts on the distribution of earnings. Oxford Economic Papers, 3, 135-46.

Rubin, D .B. (1976). Inference and missing data. Biometrica, 63, 581-592.

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. Journal of the American Statistical Association, 72, 538-543.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, NY: Wiley.

Schildhaus, S., Gerstein, D., Brittingham, A., Gray, F., Dugoni, B., Rubin, R., & Williams, E. (1998). Services research outcomes study. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Schutz, C., Rapiti, E., Vlahov, D., & Anthony, J. (1994). Suspected determinants of enrollment into detoxification and methadone maintenance treatment among injecting drug users. Drug and Alcohol Dependence, 36, 129-138.

Shah, B., Barnwell, B., Hunt, P., & LaVange, L. (1994). SUDAAN user's manual: Release 6.4. Research Triangle Park, NC: Research Triangle Institute.

Sheehan, M., Oppenheimer, E., & Taylor, C. (1988). Who comes for treatment: Drug misusers at three London agencies. British Journal of Addiction, 83, 311-320.

Simpson, D., & Curry, S. (1997). Special issue: Drug abuse treatment outcome study (DATOS). Psychology of Addictive Behaviors, 11, 211- 337.

Sobell, L., Sobell, M., & Toneatto, T. (1992). Recovery from alcohol problems without treatment. In N. Heather, W. Miller, & J. Greeley (Eds.) Self-control and the addictive behaviors (pp 198-242). New York, NY: MacMillan.

Stolzenberg, R. M., & Relles, D. (1990). Theory testing in a world of constrained research design: The significance of Heckman's censored sampling bias correction for nonexperimental research. Sociological Methods and Research, 18, 395-415.

Stolzenberg, R.M., & Relles, D. (1997). Tools for intuition about sample selection bias and its correction. American Sociological Review, 62, 494-507.

Substance Abuse and Mental Health Services Administration. (1993). National drug and alcoholism treatment unit survey (NDATUS): 1991 main findings report. Rockville, MD: Office of Applied Studies.

Substance Abuse and Mental Health Services Administration. (1997a). Uniform facility data set (UFDS): Data for 1995 and 1980-1995. Rockville, MD: Office of Applied Studies.

Substance Abuse and Mental Health Services Administration. (1997b). National household survey on drug abuse: Main findings 1995. Rockville, MD: Office of Applied Studies.

Weisberg, S. (1980). Applied linear regression. New York, NY: Wiley.

Winship, C., & Mare, R. (1992). Models for selection bias. Annual Review of Sociology, 18, 327-350.

APPENDIX

DESCRIPTION OF THE NATIONAL TREATMENT IMPROVEMENT EVALUATION STUDY AND CENTER FOR SUBSTANCE ABUSE TREATMENT DEMONSTRATIONS (1990-1992)

The National Treatment Improvement Evaluation Study (NTIES) was a national evaluation of the effectiveness of substance abuse treatment services delivered in comprehensive treatment demonstration programs supported by the Center for Substance Abuse Treatment (CSAT). The NTIES project (1992-1997) was designed and performed for CSAT by the National Opinion Research Center at the University of Chicago with assistance from Research Triangle Institute. The NTIES project collected longitudinal data between FY 1992 and FY 1995 on a purposive sample of clients in treatment programs receiving demonstration grant funding from CSAT. Client-level data were obtained at treatment intake, at treatment exit, and 12 months after treatment exit. Service delivery unit (SDU) administrative and clinician (SDU staff) data were obtained at two time points, 1 year apart.

1. THE NTIES DESIGN

1.1 The Administrative/Services Component

The NTIES study design had two levels—an administrative or services component and a clinical treatment outcomes component. The administrative component was designed to assess how CSAT demonstration funds were used, what improvements in services were implemented at the program level, and what kind and how many programs and clients were affected by the demonstration awards. Four data collection instruments were used to gather administrative/services data: the NTIES Baseline Administration Report (NBAR), the NTIES Continuing Administrative Report (NCAR), the NTIES Exit Log, and the NTIES Clinician Form (NCF).

The unit of analysis for the administrative component was the SDU, defined by CSAT as a single site offering a single level of care. The classification of *level of care* is based on three parameters:

- Facility type (e.g., hospital, etc.)
- Intensity of care (e.g., 24-hour, etc.)
- Type of service (e.g., outpatient, etc.).

An SDU could be a stand-alone treatment provider, or it could be one component of a multi-tiered treatment organization. For example, a large, county mental health agency may be the *organization* within which the SDU is located. The organization may have multiple substance abuse treatment components, such as a county hospital and a county (ambulatory) mental health center. The county hospital may have multiple SDUs, such as an inpatient detoxification service, an outpatient counseling service, and a hospital satellite center providing transitional care. In summary, the SDU provided NTIES evaluators with a stable, uniform level of comparison for examining service delivery issues.

A range of key clinician-specific data elements (within the administrative component) were assessed using the NCF. The NCF items were an important adjunct to the facility- (SDU) level instruments; these items assessed clinician training, experience, client exposure, and service provision, and were completed by all counseling and clinical (medical and therapeutic) staff at the individual SDUs.

1.2 Clinical Treatment Outcomes Component

The unit of analysis for the clinical treatment outcomes component was individual client data. NTIES measured the clinical outcomes of treatment primarily through a "before/after" or "pre- to post-treatment" design. This method compares behaviors or other individual characteristics in the same participants, measured in similar ways, before and after an intervention.

Information about clients' lives for the *before* period were obtained from the NTIES Research Intake Questionnaire (NRIQ), which was administered sometime during the clients' first 3 weeks of treatment. The specific areas assessed included:

- Drug and alcohol use
- Employment
- Criminal justice involvement and criminal behaviors
- Living arrangements
- Mental and physical health.

Information about clients' lives for the *after* period were obtained from the NTIES Post-discharge Assessment Questionnaire (NPAQ), with the same areas assessed at roughly 12 months post-treatment. Other client data sources included a treatment discharge interview (NTIES Treatment Experience Questionnaire, NTEQ), abstracted client records, urine drug screens collected at the time of the follow-up interview, and arrest reports from state databases.

1.3 The Outcome Analysis Sample

Between August 1993 and October 1994, research staff successfully enrolled 6,593 clients at 71 SDUs to participate in three waves of an in-person, computer-assisted data collection protocol. These SDUs were chosen from the universe of treatment units receiving demonstration grant funding from CSAT. Some of the selected facilities were wholly supported by CSAT awards, while others received only indirect support or none.

Clients were interviewed three times: shortly after admission on their first day of treatment, when they left treatment, and 12 months after the end of treatment. Less than 10 percent of the eligible clients refused or avoided participation, and more than 83 percent of the recruited individuals (5,388 clients) completed a follow-up interview. Additional sample exclusions included:

- Missing or undetermined treatment exit date
- Inappropriate length of follow-up interval (less than 5 or more than 16 months)
- Clients incarcerated for most or all of the follow-up period (nearly all had been treated while incarcerated, and were not yet released).

The additional sample exclusions resulted in a final outcome analysis sample of 4,411 individuals.

2. TREATMENT DEMONSTRATION PROGRAMS

CSAT initiated three major demonstration programs and made 157 multi-year treatment enhancement awards across 47 states and several territories during 1990 through 1992. One objective common to all demonstrations was CSAT's emphasis on the provision of "comprehensive treatment" services to targeted client populations. The recipients of these awards focused special attention on the substance abuse treatment service needs of minority and

special populations located primarily within large metropolitan areas. The demonstration programs are briefly described below.

2.1 Target Cities

Under this demonstration, nine metropolitan areas were selected to receive awards, of which half were included in the NTIES purposive sample. The following treatment improvement activities were explicitly provided for in the awards:

- Establishment of a Central Intake Unit (CIU) with automated client tracking and referral systems in place
- Provision of comprehensive services, including vocational, educational, biological, psychological, informational, and lifestyle components
- Improved inter-agency coordination (e.g., mental health, criminal justice, and human service agencies)
- Services for special populations—adolescents, pregnant and postpartum women, racial and ethnic minorities, and public housing residents.

2.2 Critical Populations

Under this demonstration program, awardees were required to implement "model enhancements" to existing treatment services for one or more of the following critical populations: racial and ethnic minorities, residents of public housing, and/or adolescents. Special emphasis was given to services provided to the homeless, the dually diagnosed, or persons living in rural areas. A total of 130 grants were awarded, covering services such as vocational support/counseling, housing assistance, integrated mental health and/or medical services, coordinated social services, culturally directed services, and others.

2.3 Incarcerated and Non-Incarcerated Criminal Justice Populations

Under this demonstration program, funds were directed toward improving the standard of comprehensive treatment services for criminally involved clients in correctional and other settings. Some program emphasis was placed on ethnic and/or racial minorities. Nine correctional setting demonstrations were funded: five in prisons, three in local jails, and one across a network of juvenile detention facilities. All projects included a screening component to

identify substance-abusing inmates, a variety of targeted treatment interventions (e.g., therapeutic communities, intensive day treatment programs), and a substantial aftercare component.

A total of 10 non-incarcerated projects were funded. Five programs targeted interventions at clients in diversionary programs, three focused services on probationers or parolees, and two programs targeted both populations. Almost all of the funded demonstration projects included the following components:

- Basic eligibility determination, followed by systematic screening and assessment
- Referral to treatment
- Graduated sanctions and incentives while in treatment
- Intensive supervision in treatment
- Community-based aftercare with supervision and service coordination.

In total, 19 criminal justice projects were funded as part of the CSAT 1990-1992 demonstrations, and as indicated in the next section, these projects were purposively over-sampled in order to obtain a more robust evaluation of this program.

3. DESCRIPTION OF SDUS AND CLIENTS BY TREATMENT MODALITY AND PROGRAM TYPE

The 71 SDUs contributing clients to the outcome analysis sample are characterized by modality and (demonstration) program type in Exhibit A-1 below. Among the 698 SDUs in the NTIES universe: 52 percent (n=365) were Target Cities programs, 39 percent (n=274) were Critical Populations programs, and 9 percent (n=59) were Criminal Justice programs.

In terms of the SDUs sampled for the NTIES outcome analysis, 44 percent were Target Cities programs, 38 percent were Critical Populations programs, and 23 percent were Criminal Justice programs. Criminal Justice SDUs were purposely over-sampled as part of the NTIES evaluation design (CSAT, 1997). Nearly half of the sampled SDUs were (non-methadone) outpatient programs, and about one-quarter were long-term residential programs.

EXHIBIT A-1
SDUs IN THE OUTCOME ANALYSIS SAMPLE

Program Title Number of SDUs (% of NTIES Universe)[1]	NTIES Sample	Methadone	Outpatient	Long-Term Residential	Short-Term Residential	Correctional
Target Cities n=365 (52%)	31 (44%)	6	15	6	4	0
Critical Populations n=274 (39%)	27 (38%)	1	13	10	3	0
Criminal Justice n=59 (9%)	13 (23%)	0	5	0	0	8
Totals N=698 (100%)	71 (100%)	7	33	16	7	8

EXHIBIT A-2
DISTRIBUTION OF CLIENTS IN THE OUTCOMES ANALYSIS SAMPLE

Program Title Number of Clients (% of Analysis Sample)	Methadone	Outpatient	Long-Term Residential	Short-Term Residential	Correctional
Target Cities n=2,600 (59%)	377 (89%)	1,214 (78%)	504 (60%)	505 (58%)	0
Critical Populations n=931 (21%)	45 (11%)	220 (14%)	298 (35%)	368 (42%)	0
Criminal Justice n=880 (20%)	0	132 (8%)	39 (5%)	0	709 (100%)
Totals n=4,411 (100%)	422	1,566	841	873	709

BEST COPY AVAILABLE

[1] The original NTIES universe of SDUs included a program type called *Specialized Services*. Because clients for the outcome analysis sample were not drawn from these SDUs (n=94), they are excluded from the Exhibit.

As shown in Exhibit A-2, 59 percent of all NTIES clients were sampled from Target Cities SDUs. Slightly over 21 percent of all NTIES clients were sampled from Critical Populations SDUs, and 20 percent were sampled from Criminal Justice SDUs. Outpatient (non-methadone) SDUs treated over one-third (35%) of the clients in the outcomes analysis sample, and almost 80 percent of these were sampled from Target Cities programs.

Readers who are interested in more detailed information about the NTIES project are invited to visit the NEDS Web site at: http://neds.calib.com. The NEDS Web site provides the full-length version of the NTIES Final Report (1997), as well as copies of all data collection instruments employed in NTIES.